

Bachelorarbeit

von Klara Zeis

Sommersemester 2013

Was sind die Risikofaktoren für Rehe, vom Luchs gerissen zu werden?

Analysen unter Verwendung von Cox - Modellen mit zeitvariierenden Kovariablen.

Betreuung:

Prof. Dr. Helmut Küchenhoff

Institut für Statistik

Ludwig-Maximilians-Universität München



Im Zusammenarbeit mit dem Nationalpark Bayerischer Wald

Inhaltsverzeichnis

1. Einleitung	1
1.2 Das Reh	1
1.3 Ziel der Arbeit	1
2. Daten	2
2.1 Rohdaten	2
2.1.1 Aktivdaten	3
2.1.2 Kopfdaten	4
2.2 Kodierung und Bereinigung der Daten	4
2.2.1 Aktivwertbestimmung bzgl. der Tageszeiten	5
2.2.2 Aktivanteil	6
2.2.2 Bereinigung des Rehdatensatzes	7
2.2.3 Kodierung des Alters der Rehe	9
2.2.4 Kodierung des WSI pro Monat	9
2.2.5 Kodierung der Variable „Habitat“	10
2.2.6 Kodierung des Beobachtungszeitraumes	11
3. Deskriptive Analysen	13
3.1. Aktivität	13
3.2 Aktivanteil	14
3.3 Luchsriss	15
3.3.1 Luchsriss und Alter	16
3.3.2 Luchsriss und Geschlecht	17
3.3.3 Aktivitätswerte und Luchsriss	18
3.3.4 Aktivanteile und Luchsriss	19
4. Das Cox-Modell	20
4.1 Theorie zum Cox-Modell mit festen Kovariablen	20
4.1.1 Die Survivalfunktion und Kaplan-Meier-Schätzer	20
4.1.2 Die Hazard-Funktion	21
4.1.3 Das Cox-Modell mit festen Kovariablen	21
4.1.4 Hazard Ratio	23
4.1.5 Bindungen	23
4.2 Theorie zum Cox-Modell mit zeitvariierenden Kovariablen	24
4.2.1 Zählprozess und Zeitskalen	25
4.2.2 Das Cox-Modell mit zeitvariierenden Kovariablen	26
4.3 Residuen und Test auf PH-Annahme	27

4.3.1 Cox-Snell-Residuen	27
4.3.2 Martingale Residuen	28
4.3.3 Grambsch und Therneau Test	28
5. Datenanalyse des Rehdatensatzes	30
5.1 Korrelationen zwischen den Kovariablen	30
5.2 Cox-Modell mit festen Kovariablen.....	33
5.2.1 Modell 1f.....	34
5.2.1 Residuen Modell 1f.....	36
5.2.2 Test auf PH-Annahme	37
5.2.3 Modell 1f mit pspline.....	38
5.3 Erweiterung des Cox-Modells mit zeitvariierenden Kovariablen	39
5.4 Zeitvariierendes Cox-Modell mit Aktivwerten	41
5.4.1 Modell 1	41
5.4.2 Modell 2	43
5.5 Zeitvariierendes Cox-Modell mit Aktivanteilen	45
5.5.1 Modell 1a.....	45
5.5.2 Modell 2a.....	47
5.5.3 Modell 3a.....	49
5.6 Cox-Snell Residuen der Modelle mit variierenden Kovariablen	51
5.7 Martingal-Residuen	52
5.8 Test auf PH-Annahme	53
6. Fazit.....	54
Literaturverzeichnis.....	56
Abbildungsverzeichnis	57
Eidesstattliche Erklärung	59
Danksagung	60
Anhang.....	61

1. Einleitung

Im Nationalpark Bayerischer Wald werden seit 2005 die Wechselwirkungen zwischen Luchs und deren Beutetieren Reh und Rothirsch untersucht. Der Luchs ist das größte, katzenartige Raubtier und neben Bär und Wolf der drittgrößte Beutegreifer Europas. In den 80er Jahren wurden 17 dieser Tiere im tschechischen Nationalpark Šumava ausgewildert. Diese Tiere bildeten den Grundstock der heutigen Luchspopulation im Nationalpark bayerischer Wald und Šumava. Nach dem Rückgang der Luchspopulation Ende der 90er Jahre wurde ein Luchsprojekt in Leben gerufen, das die Ökologie des Luchses und seines Hauptbeutetiers Reh genau untersuchen soll. (Leibl, 2012)

1.2 Das Reh

In der Bachelorarbeit werden die Risikofaktoren der Rehe bzgl. eines Luchsrisses mittels erweiterter Survivalanalysen untersucht. In diesem Kapitel wird nun erst einmal auf die Lebensweise und Biologie der Rehe eingegangen.

Aufgrund seiner sehr guten Anpassung an verschiedenste Habitate ist das Reh (*Capreolus capreolus*) in Europa sehr weit verbreitet. Die natürlichen Lebensräume der Rehe reichen von Laub-, Nadel- und mediterranen Wäldern bis hin zu Mooren und Sümpfen. Was seine Art aber so erfolgreich macht ist die Anpassung an den Menschen. So lebt das Reh in den forstwirtschaftlich genutzten Wäldern, in landwirtschaftlich genutzten Feldern, Wiesen und Äckern. Dabei ernährt sich das Reh von pflanzlichen Kost und Wurzeln. (Andersen, et al., 1998). Der Sexualzyklus der Rehe ist bestimmt zwischen dem inneren Rhythmus und der Tageslänge. Rehe verfügen über eine sog. Keimruhe, das bedeutet, dass sich das während der Paarungszeit befruchtete Ei erst ab Dezember entwickelt und somit die Geburt in den Frühling fällt. Das Ergebnis dieses gesteuerten Sexualrhythmus ist, dass in Mitteleuropa 80% der Jungtiere im Mai geboren werden (Wahlström, et al., 1998).

1.3 Ziel der Arbeit

In dieser Bachelorarbeit sollen anhand einer Survivalanalyse unter Verwendung des Cox-Modells mit zeitvariierender Kovariablen geklärt werden, welche Risikofaktoren der Rehe die Wahrscheinlichkeit eines Luchsrisses erhöhen. Dabei werden die Datenkodierungen und Analysen des vorangegangenen statistischen Praktikums „Analyse des Einflusses zwischen

individuellem Verhalten der Rehe und der Wahrscheinlichkeit eines Luchsrisses“ (Zeis, et al., 2013) übernommen und erweitert. Die Risikofaktoren konzentrieren sich weiterhin auf die Aktivität der Tiere. Allerdings soll hier nicht nur die Aktivität in der nautischen Dämmerung untersucht werden, sondern auch die Aktivität in der Nacht und am Tag. Ein weiteres Merkmal für die Untersuchung wird der Aktivanteil sein, dieser wird ebenfalls in Dämmerung, Tag und Nacht aufgeteilt und als Variable aufgenommen. Zusätzlich werden die Variablen Geschlecht, Alter, WSI und Habitat in die Untersuchung mit einbezogen.

2. Daten

Die untersuchten Rehe wurden im Bayerischen Nationalpark mittels einer Kastenfalle gefangen und anschließend mit einem Halsband besendert. Die Halsbänder stammen von der Firma *Vetronic Aerospace GmbH*, die sich auf *Wildlife Monitoring* spezialisiert hat und zu den weltweit führenden Firmen in diesem Bereich gehört. Die Halsbänder liefern Daten zur GPS-Bestimmung, die die Position der Tiere wiedergibt und per SMS an den Anwender weitergibt. Sie enthalten außerdem Daten zur Aktivitätsmessung und Temperatur (Butz, et al., 2010). Die Aktivdaten, siehe Kapitel 2.1.1 „Aktivdaten“, werden erst nach der Abnahme des Halsbandes ausgewertet. Um mit den Daten im erweiterten Cox-Modell arbeiten zu können müssen diese kodiert und bereinigt werden. In diesem Kapitel wird nun auf die Kodierung und Bereinigung der Daten genauer eingegangen.

2.1 Rohdaten

Die Daten der Untersuchung stammen von den ausgewerteten Halsbandsendern der oben genannten Firma. Sie enthalten Messungen über einem Zeitraum von 2004 bis 2012. Insgesamt wurden von dem Projektpartner 185 sogenannte Aktivdaten auf einer CD und 232 sogenannte Kopfdaten in einer Tabelle zur Verfügung gestellt. Die Daten in dieser Untersuchung enthalten also keine GPS-Daten. Allerdings wird, aufgrund einer Zusammenarbeit mit einer anderen Bachelorarbeit, die sich mit dem Thema „*Habitatwahl und Risiko für Rehe unter Luchsprädatoren*“ beschäftigt, eine Variable namens „Habitat“ aufgenommen. Diese Variable wird mit GPS-Daten berechnet.

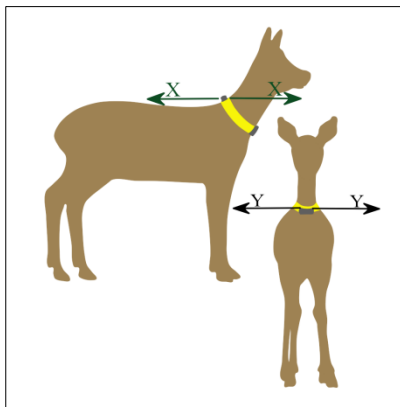
2.1.1 Aktivdaten

Die übergebenen Aktivdaten enthielten die folgenden Variablen:

- „*utc*“, „*act_x*“, „*act_y*“, „*temp*“

Dabei gibt „*utc*“ die koordinierte Weltzeit an, diese entspricht der Mitteleuropäischen Zeit (MEZ) + 1 Stunde in der Winterzeit und MEZ + 2 Stunden in der Sommerzeit.

Die Variablen „*act_x*“ und „*act_y*“ geben die skalierten Bewegungen der Tiere wieder. Die Variable „*act_x*“ misst hierbei die Vor- und Rückwärtsbewegungen, während „*act_y*“ die



seitlichen Bewegungen misst. Beide Bewegungsdaten haben einen Wertebereich von 0 bis 255.

Die Messungen der Variable „*temp*“ zeigen die Temperatur des Senders an, sie messen somit eine Mischung aus Außentemperatur und Temperatur des Tieres. Sämtliche Variablen wurden in 5-Minuten-Schritten aufgezeichnet. (Butz, et al., 2010)

Abbildung 1: Richtungsmessungen

Tabelle 1 zeigt die ersten fünf Zeilen der Aktivdaten des Rehes „Theresa“ am 16.12.2009 von 7:50 Uhr bis 8:10 Uhr. In den ersten zehn Minuten hat das Tier geruht, daher wurde keine Bewegung gemessen und beide Achsen haben den Wert 0. Ab 8:00 Uhr wurden die ersten Bewegungen des Tieres aufgezeichnet. Dabei sind die Werte in der Tabelle die absoluten Aktivwerte, die noch auf einen Aktivwert berechnet werden, siehe dazu Kapitel 2.2.1 „Aktivwertbestimmung“. Der Anteil zwischen der Aktivität und der Ruhephase wird ebenfalls als eigene Variable berechnet und in das Modell aufgenommen, siehe dazu Kapitel 2.2.2 „Aktivanteil“.

utc	act_x	act_y	temp
2009-12-16 07:50:00	0	0	9
2009-12-16 07:55:00	0	0	9
2009-12-16 08:00:00	1	1	9
2009-12-16 08:05:00	72	95	8
2009-12-16 08:10:00	30	39	7

Tabelle 1: Erste Zeilen der Rohdaten des Rehes "Theresa" vom 16.12.2009

2.1.2 Kopfdaten

Die 175 „Kopfdaten“ der Rehe wurden in Tabellenform übergeben. Jedes Reh hat einen Namen und eine Halsbandnummer, einige Rehe wurden wiederbesendert, das bedeutet, dass ein Reh verschiedene Halsbänder getragen hat. Deshalb befinden sie sich mehrmals in der Tabelle, jedoch mit unterschiedlichen Halsbandnummern.

Außerdem enthalten die Kopfdaten noch Informationen über den Sender (die Nummer des Halsbandes, welches das Reh getragen hat), die Auslesesoftware, die Messzeitraumangaben „*Signal_on*“ und „*Signal_off*“, sowie Angaben über das geschätzte Alter bei der Besenderung der Tiere und das Schicksal des Rehes.

Tabelle 2 zeigt einen Ausschnitt der ersten Zeilen der sog „Kopfdaten“. In der ersten Zeile werden die Daten des Rehs „Aischa“ angezeigt, es wurde vom 30.03.2011 bis zum 09.01.2012 beobachtet, das Reh trug ein Halsband mit dem Sender 9372, dieser wurde mit der Software 2.7.3 ausgelesen. Bei der Besenderung wurde das Alter des Tieres auf unter ein Jahr geschätzt und durch die Spalte „Schicksal“ wird wiedergegeben, dass Reh „Aischa“ von einem Luchs getötet wurde.

Name	Sender	Signal_on	Signal_off	Species	Age	Schicksal	Software
Aischa	9372	30.03.2011	09.01.2012	Reh	<1	tot: gerissen	2.7.3
Andrea	7621	11.01.2012	17.01.2012	Reh	3	tot: gerissen	2.7.7
Anna	2315	13.02.2008	11.12.2008	Reh	1	tot	1.7.3
Annika	7621	26.01.2012	24.02.2012	Reh	3	tot: gerissen	2.7.7

Tabelle 2: Ausschnitt aus den ersten Zeilen der sog "Kopfdaten"

2.2 Kodierung und Bereinigung der Daten

Da die Aktivdaten in 5-Minuten-Schritten aufgenommen wurden, ergab sich für alle Datensätze der untersuchten Rehe eine Größe von 430 MB. Damit das Statistikprogramm „R“ mit dieser Anzahl von Daten umgehen konnte, wurde jeder Datensatz einzeln bereinigt, kodiert und anschließend zu einer großen Tabelle zusammengefasst.

2.2.1 Aktivwertbestimmung bzgl. der Tageszeiten

In dem statistischen Vorgängerpraktikum wurde auf Wunsch des Projektpartners für jedes Reh ein Aktivitätswert in den nautischen Dämmerungszeiten, da hier auch die Jagdzeit des Luchses liegt, bestimmt (Zeis, et al., 2013). In dieser Arbeit soll nun zusätzlich eine Aktivitätsbestimmung am Tag und in der Nacht erfolgen. Außerdem wird die Aktivitätszahl für das Cox-Modell mit zeitvariierenden Kovariablen nicht mehr insgesamt, sondern als zeitvariierende Kovariable pro Monat bestimmt.

Aus den Ergebnissen des Vorgängerprojekts ergab sich eine starke Korrelation zwischen den „*act_x*“- und den „*act_y*“-Werten. Der Korrelationskoeffizient nach Pearson betrug 0,957 (Zeis, et al., 2013). Beide Werte wurden daraufhin zusammengezählt und durch die Anzahl der Beobachtungen geteilt, um so einen einzigen Aktivwert zu erhalten.

Damit die Aktivität in der Dämmerung, sowie am Tag und in der Nacht bestimmt werden konnte, wurden Daten der NASA, aus dem Internet (Galupki, 2012) bezogen. Diese Daten, die Werte der Tageslänge, Sonnenuntergang und Sonnenaufgang enthalten, beziehen sich auf die Ortschaft Riedlhütte im Bayerischen Wald. Die Nasadaten geben weiterhin an, wann an einem betreffenden Tag der nautische und offizielle Sonnenauf- und Sonnenuntergang stattgefunden hat, so lassen sich die Dämmerungszeiten auf die Minute genau bestimmen.

Die nautische Morgendämmerung beginnt mit dem nautischen Sonnenaufgang und endet mit dem offiziellen Sonnenaufgang. Die nautische Abenddämmerung beginnt mit dem offiziellen Sonnenuntergang und endet mit dem nautischen Sonnenuntergang.

Vor dem Zusammenfügen der Daten wurde die Sommerzeit aus den Nasadaten herausgerechnet und die koordinierte Weltzeit auf die Mitteleuropäische Zeit umgestellt. Um einen Aktivwert pro Monat und Tageszeit zu bestimmen wurde folgendermaßen vorgegangen: Erst wurde der Datensatz pro Reh in einzelne Monate aufgeteilt, die Daten der Monate wurden nochmals in die einzelnen Tageszeiten (nautische Dämmerung, Tag und Nacht) gesplittet. Anschließend wurden alle *act_x* und *act_y* – Werte aufsummiert und durch die Anzahl der Beobachtungen geteilt. So wurde je ein Aktivwert der Dämmerung, des Tages und der Nacht pro Reh und pro Monat erstellt.

Abbildung 2, die anhand der Nasadaten (Galupki, 2012) angefertigt wurde, zeigt den Verlauf der Dämmerungs-, Tages- und Nachtzeiten innerhalb eines Jahres. Dabei ist der orange Bereich der Zeitraum der nautischen Dämmerung, der grüne Bereich der Zeitraum des Tages und der blaue Bereich der Zeitraum der Nacht. Deutlich zu erkennen ist die Verschiebung der Dämmerungszeiten, die sich täglich um einige Minuten ändern. So ist die Dämmerung in den Sommermonaten länger und tritt früher am Morgen und später am Abend auf. In den Wintermonaten dagegen tritt die Dämmerung später am Morgen und früher am Abend auf und dauert insgesamt kürzer als in den Sommermonaten.

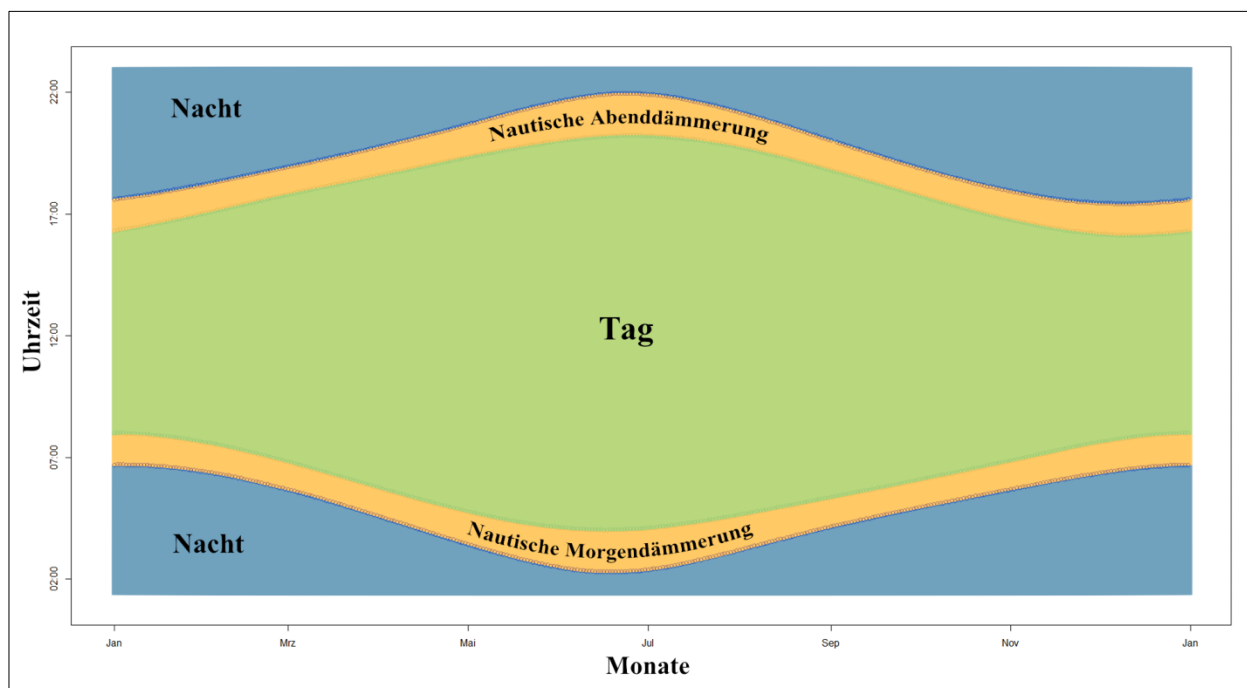


Abbildung 2: Darstellung der Aufteilung der Tageszeiten im Jahresverlauf

2.2.2 Aktivanteil

Im Gegensatz zu den Aktivwerten, die die gemittelte Höhe der absoluten Aktivitäten angeben, misst die Variable „Aktivanteil“ den Anteil zwischen aktivem und ruhendem Verhalten pro Monat. Dazu wird jede einzelne Beobachtung entweder mit 1 oder 0 kodiert, je nachdem ob bei einem Tier eine Bewegung ≥ 1 aufgezeichnet wurde oder keine Bewegung gemessen wurde. Werden nun die mit 1 kodierten Daten aufsummiert und durch die gesamte Anzahl der Beobachtungen geteilt, ergibt sich ein Anteilswert zwischen 0 und 1. Mit diesem Anteilswert kann nun prozentual angegeben werden wie aktiv ein Reh sich in diesem Monat verhalten hat.

In Abbildung 3 sieht man beispielhaft die Aktivität von Reh „David“ am 07.03.2010. Die orange Linie zeigt hier den tatsächlichen Messverlauf der absoluten Aktivwerte, die in fünf Minuten Intervallen aufgezeichnet wurde. Die blauen Punkte geben an, ob eine Aktivität (1) oder eine Ruhephase (0) vorlag. Bei diesem Tag gab es 145 Intervalle in denen eine Aktivität gemessen worden ist und 139 Intervalle in denen keine Aktivität, somit eine Ruhephase gemessen wurde. Insgesamt errechnet sich somit ein Anteilswert von 0.51, was einen Prozentwert von 51 % entspricht. Der Aktivanteil des Rehes lag also an diesem Tag bei 51%. Die Abbildung zeigt allerdings die Aktivwerte bzw. den Aktivanteil eines gesamten Tages, in der Kodierung der Daten wurde hier die Trennung der Tageszeiten berücksichtigt. Daraus folgt, dass es pro Tag drei Aktivanteile gibt, einen in der Dämmerung, einen am Tag und einen in der Nacht.

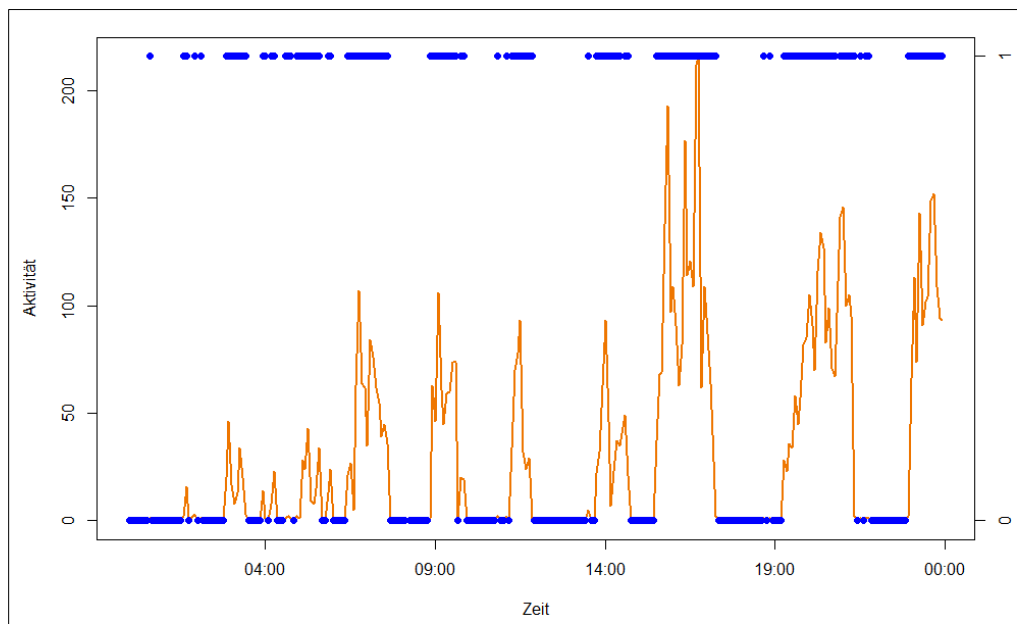


Abbildung 3: Aktivwerte von Reh "David" am 07.03.2010

2.2.2 Bereinigung des Rehdatensatzes

Wie schon im zweiten Kapitel beschrieben werden die Aktivdaten nicht sofort an den Anwender weitergegeben, sondern erst nach der Abnahme des Halsbandes mittels eines sog „Link Managers“ per USB-Anschluss mit dem PC ausgelesen. (Butz, 2010)

Die Software, die zum Auslesen dieser Daten verwendet wurde, ändert sich, aus Update Gründen, kontinuierlich. In der Untersuchung des statistischen Vorpraktikums zeigten sich einige starke Ausreißer bei den Aktivdaten. Die Ausreißer waren alle ausschließlich auf die

Software 2.0.8 zurückzuführen. Man muss nun davon ausgehen, dass diese Software fehlerhaft ist. Aus diesen Gründen wurden die Daten, die mit dieser Software ausgelesen wurden aus dem Datensatz entfernt. (Zeis, et al., 2013)

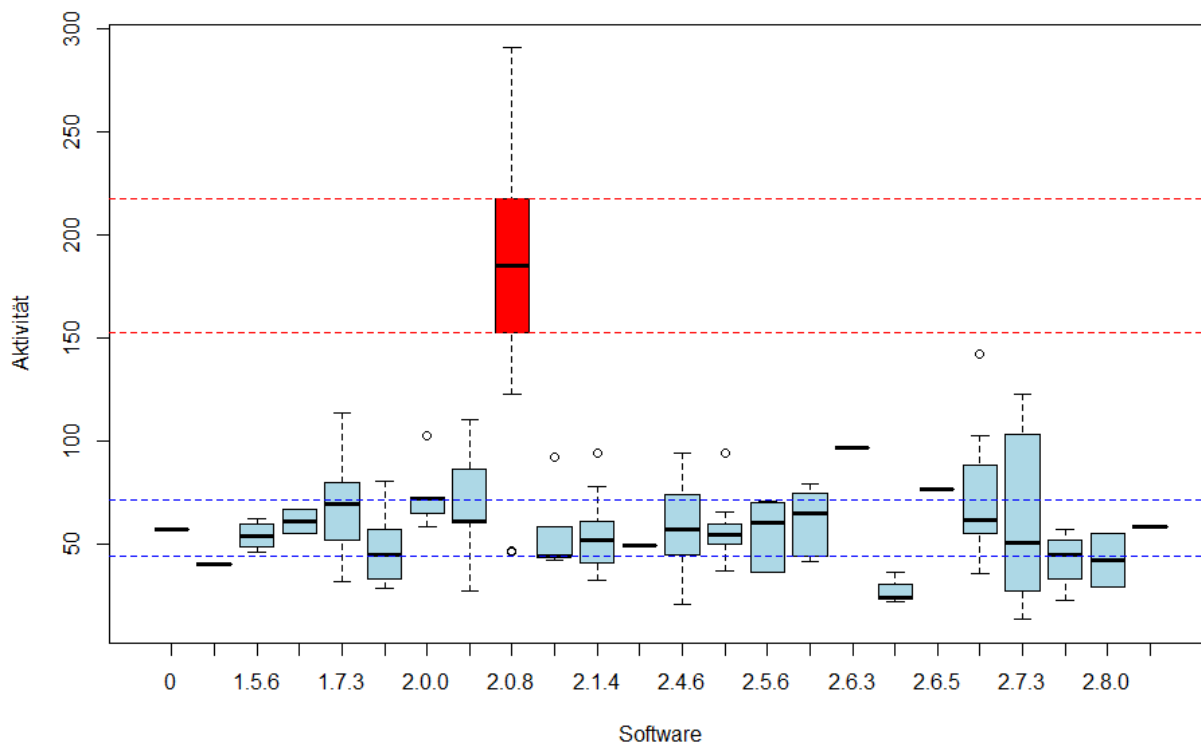


Abbildung 4: Fehlerhafte Software 2.0.8

Wie in Abbildung 4 zu erkennen ist, werden mit Software 2.0.8 deutlich höhere Aktivdaten ausgelesen, als mit der übrigen Software. Dabei zeigt die X-Achse die verwendete Software und die Y-Achse die Aktivität, die gemessen wurde. Die gemessene Aktivität ist der Aktivwert, der bereits in Kapitel 2.2.1 beschrieben wurde. 50% der mit der fehlerhaften Software ausgelesenen Daten liegen zwischen 152 und 217, während 50% der übrigen Daten zwischen 43 und 72 liegen. Der Median in Software 2.0.8 beträgt 185, der Median in den übrigen Softwaretypen bei 56. Von den 13 Rehen, deren Halsbänder mit der fehlerhaften Software ausgelesen wurden, sind 6 Tiere wiederbesendert worden. Diese Daten konnten weiterhin in die Untersuchung analysiert werden, somit wurden durch die fehlerhafte Software insgesamt 7 Rehe aus der Untersuchung entfernt.

Durch fehlende „Kopfdaten“, sowie Nullmessungen mussten weiterhin 8 Tiere aus der Untersuchung ausgeschlossen werden. Nullmessungen sind Messwerte, die keinerlei Aktivität aufgenommen haben. Insgesamt blieben so 120 Rehe in der Enduntersuchung.

2.2.3 Kodierung des Alters der Rehe

Das Alter der Rehe wurde in den Kopfdaten übermittelt. Es handelte sich dabei um das geschätzte Alter bei Besenderung. Das Reh wird mit einem Alter unter 12 Monaten als Kitz, zwischen 12 und 24 Monaten als „Schmalreh“ (weibliches Tier) oder „Jährling“ (männliches Tier) und ab einem Alter von 24 Monaten als „Geiß“ (weibliches Tier) oder „Bock“ (männliches Tier) bezeichnet.

Um die Altersangaben als Kovariable aufnehmen zu können, wurde das Alter bei Besenderung auf das geschätzte Geburtsjahr zurückgerechnet und schließlich auf das Alter in Monaten umgerechnet. Dabei gilt das Alter als feste Größe und bezeichnet das Alter des Tieres zu Studienbeginn. Anzumerken wäre hierbei noch, dass als Geburtsmonat immer der Mai verwendet wurde, da in dieser Zeit 80% der Kitze geboren werden (Wahlström, et al., 1998).

2.2.4 Kodierung des WSI pro Monat

WSI bezeichnet den „Winter Streng Index“. Dabei handelt es sich um einen Index, der die Härte eines Winters bestimmt. Hierzu werden die Tage in den Wintermonaten Dezember, Januar, Februar, März und April gezählt, bei denen die Temperatur $\leq 0^{\circ}\text{F}$ ($-17,76^{\circ}\text{C}$) und eine Schneehöhe von ≥ 18 Inches (45,72 cm) beträgt. Jeder Winter bekommt so einen Index, der anschließend in *mild* (weniger als 50 Tage), *normal* (50 – 80 Tage) und *streng* (mehr als 80 Tage) eingeteilt wird.

Die Daten zum WSI wurden anhand einer Tabelle vom bayerischen Nationalpark übergeben. In dieser Tabelle waren für jeden Wintermonat von 1970 – 2012 die Anzahl der betreffenden Tage aufgelistet. Für eine Untersuchung mit zeitvariierenden Kovariablen wurde, nach Absprache mit dem Projektpartner, der monatliche Wert aufgenommen und nicht mehr der WSI des Winters insgesamt. Es wäre in unserem Fall also von einem Wintermonat-Streng-Index zu sprechen. Die restlichen Monate wurden mit 0 kodiert. Auf eine Einteilung in „mild“, „mittel“ und „streng“ wurde ebenfalls verzichtet. Da in einem Cox-Modell keine zukünftigen Ereignisse aufgenommen werden dürfen, siehe dazu Kapitel 4.2.2 „Das Cox-Modell mit zeitvariierenden Kovariablen“, wurde der WSI des jeweiligen Vormonats aufgenommen.

Ein Überblick über die in der Untersuchung fallenden Monate und ihrer Winterhärte gibt Tabelle 3 wieder. Dabei ist zu beachten, dass im März 2012 eine sehr strenge Winterhärte mit einem Wert von 34 herrschte, während z.B. in den gesamten Wintermonaten 2006/2007 kein einziger Tag verzeichnet wurde, an dem die Schneehöhe ≥ 18 Inches oder eine Temperatur $\leq 0^\circ\text{F}$ vorlag.

Klimastation Waldhäuser		Dez	Jan	Feb	Mrz	Apr	WSI
2004/05	Schneehöhe $\geq 45,72\text{cm}$	0	12	28	28	0	68
	$T_{\min} \leq -17,78^\circ\text{C}$	0	0	0	0	0	
2005/06	Schneehöhe $\geq 45,72\text{cm}$	16	31	28	31	14	120
	$T_{\min} \leq -17,78^\circ\text{C}$	0	0	0	0	0	
2006/07	Schneehöhe $\geq 45,72\text{cm}$	0	0	0	0	0	0
	$T_{\min} \leq -17,78^\circ\text{C}$	0	0	0	0	0	
2007/08	Schneehöhe $\geq 45,72\text{cm}$	9	19	3	7	0	38
	$T_{\min} \leq -17,78^\circ\text{C}$	0	0	0	0	0	
2008/09	Schneehöhe $\geq 45,72\text{cm}$	0	0	18	31	0	49
	$T_{\min} \leq -17,78^\circ\text{C}$	0	0	0	0	0	
2009/10	Schneehöhe $\geq 45,72\text{cm}$	0	3	24	6	0	34
	$T_{\min} \leq -17,78^\circ\text{C}$	1	0	0	0	0	
2010/11	Schneehöhe $\geq 45,72\text{cm}$	11	4	0	0	0	15
	$T_{\min} \leq -17,78^\circ\text{C}$	0	0	0	0	0	
2011/12	Schneehöhe $\geq 45,72\text{cm}$	13	31	29	23	0	101
	$T_{\min} \leq -17,78^\circ\text{C}$	0	0	5	0	0	

Tabelle 3: Winter-Streng-Index der Klimastation Waldhäuser

2.2.5 Kodierung der Variable „Habitat“

Bei jedem Reh wurde mehrmals pro Tag der Aufenthaltsort mittels GPS bestimmt. Um zu ermitteln, in welchem Habitat sich das Reh aufgehalten hat, wurde die Karte des Bayerisch-Böhmischen Nationalparks in folgende Habitate eingeteilt:

Straße, Nadelbestand-mittel, Nadelbestand-alt, Nadelbestand-jung, Mischbestand-mittel, Mischbestand-alt, Mischbestand-jung, Laubbestand-mittel, Laubbestand-alt, Laubbestand-jung, Totholz-liegend-Laubbaumverjüngung, Totholz-liegend-Nadel-Laubbaumverjüngung, Totholz-liegend, Totholz-liegend-Nadelbaumverjüngung, Totholz-stehend, Wiese-nicht bewirtschaftet, Ökoton (Mischung zwischen Wald und Wiese), Feuchtwiese, Kahlschlag, Wiese-bewirtschaftet, Moor, Felsen, Latsche, Wasserfläche, Siedlung und Eisenbahn

Schließlich konnte mittels der GPS Koordinaten das Habitat jeder Peilung herausgefunden werden. Für das Überlebenszeitmodell wurde für jedes Reh der häufigste Aufenthaltsort pro Monat ermittelt. Dazu wurden die Anzahlen der Peilungen in den Habitaten für jeden Monat aufsummiert und durch die Flächengröße des jeweiligen Habitats vom gesamten Nationalpark geteilt. Als häufigster Aufenthaltsort wurde dann das Habitat mit dem größten Wert verwendet. Mit der Flächengewichtung sollte vermieden werden, dass große Habitate stärker ins Gewicht fallen (Mayer, 2013).

2.2.6 Kodierung des Beobachtungszeitraumes

Für jedes Reh wurde im Datensatz pro Beobachtung in einem Monat eine Zeile mit dessen Werten angelegt. Wird nun z.B. ein Reh in einem bestimmten Monat an 17 Tagen beobachtet (ca. 4896 einzelne Aktivdaten, denn die Rehe werden ja in fünf Minuten Intervallen beobachtet), so wird mittels der Formel Aktivwertberechnung aus Kapitel 2.2.1 „Aktivwertbestimmung“ ein einziger Wert für den Monat Mai berechnet.

Die Variable „Beobachtungszeitraum“ gibt nun an, wie viele Monate ein Reh in der Untersuchung insgesamt beobachtet wurde. Da einige Rehe, wie bereits im Kapitel 2.2.2 „Bereinigung des Rehdatensatzes“ erwähnt, Lücken in der Beobachtung aufwiesen, gibt der Beobachtungszeitraum nur die monatliche Intervalle wieder, in denen ein Reh tatsächlich beobachtet und somit Aktivdaten übermittelt wurden. Die Intervalle, in denen keine Daten vorliegen können entweder aus dem Datensatz gelöscht werden, oder als „NA“ gekennzeichnet werden. Diese fehlenden Beobachtungen werden dann in der Auswertung von dem Statistikprogramm „R“ selbstständig aus der Datenanalyse entfernt (Therneau, et al., 2001). In dieser Arbeit sind die Intervalle, in denen keine Daten erhoben wurden, gelöscht worden. Von den 120 beobachteten Tieren weisen 11 Rehe Lücken in den Intervallen auf, der grafische Überblick der Lücken findet sich im Anhang wieder.

Am Ende der Beobachtungsintervalle wurde das Reh entweder als „zensiert“ oder als „event“ gekennzeichnet. Je nachdem, ob ein Reh vom Luchs gerissen worden ist oder nicht. Als „zensiert“ wurden die Tiere gekennzeichnet, die entweder den Tod durch ein anderes Ereignis gefunden haben oder am Ende der Studie noch gelebt haben und entsendert wurden.

Um im späteren Verlauf der Untersuchungen das Cox-Modell mit den zeitvariierenden Kovariablen berechnen zu können wurde zusätzlich jedem Reh eine Variable „start“ und „stop“ hinzugefügt, diese geben die Zeitintervalle an, in denen die Werte gemessen wurden, siehe dazu Kapitel 4.2.1 „Zählprozess und Zeitskalen“.

Nachdem sämtliche Kodierungen vorgenommen worden sind, stellt sich die Tabelle für die Untersuchung unter Verwendung des Cox-Modells mit zeitvariierenden Kovariablen folgendermaßen dar:

Name	G	start	stop	Aktivwert Däm.	Aktivwert Tag	Aktivwert Nacht	Aktivanteil Däm.	Aktivanteil Tag	Aktivanteil Nacht	WSI	Alter	Habitat	Event
Aischa	w	0	1	40,21	24,2	49,37	46,15	40,48	31,93	0	10	Nadel	0
Aischa	w	1	2	59,16	48,17	77,11	47,18	61,15	41,29	0	10	Laub	0
Aischa	w	2	3	57,01	69,22	75,88	47,55	63,43	55,35	0	10	Totholz	0
Aischa	w	3	4	52,23	60,9	64,92	51,19	60,36	56	0	10	Totholz	0
Aischa	w	6	7	50,03	44,82	51,94	50,76	56,02	46,08	0	10	Totholz	0
Aischa	w	7	8	35,34	49,97	76,47	36,41	53,99	46,06	0	10	Laub	0
Aischa	w	8	9	69,51	46,69	73,46	59,83	65,9	49,54	0	10	Laub	0
Aischa	w	9	10	78,3	43,79	81,56	57,00	53,82	41,9	0	10	Laub	0
Aischa	w	10	11	35,64	29,6	40,69	28,99	32,9	28,31	13	10	Nadel	1
Andrea	w	0	1	43,58	22,41	38,86	59,9	39,97	30,56	13	44	Nadel	1
Annika	w	0	1	37,3	17,55	36,54	50,89	46,14	28,25	19	44	Laub	0
Annika	w	1	2	19,16	18,08	41,72	28,47	51,46	24,77	31	44	Laub	1
Arabiata	w	0	1	13,09	12,44	29,85	16,00	26,98	16,56	0	10	Nadel	0
Arnold	m	0	1	43,35	23,86	46,06	54,44	48,81	35,99	13	8	Laub	0
Arnold	m	1	2	29,69	18,4	40,14	37,02	44,77	26,16	31	8	Laub	1

Tabelle 4: Ausschnitt aus der Endtabelle nach Kodierung und Bereinigung

3. Deskriptive Analysen

Für einen ersten Überblick über die Daten und die Zusammenhänge werden im folgenden Kapitel deskriptive Analysen durchgeführt. Dabei wird zuerst auf den Aktivwert und den Aktivanteil eingegangen und schließlich die anderen Variablen bzgl. eines Luchsrisses dargestellt. Da es sich nur um erste Analysen handelt, wird auf Tests verzichtet.

3.1. Aktivität

Wie bereits im Kapitel 2.2.1 „Aktivwertbestimmung bzgl. der Tageszeiten“ besprochen, wurde die Aktivität der Dämmerung, die Aktivität des Tages und die Aktivität der Nacht anhand der Beobachtungen pro Monat bestimmt. Anhand der 1324 monatlichen Beobachtungswerte wurden die folgenden Boxplots erstellt.

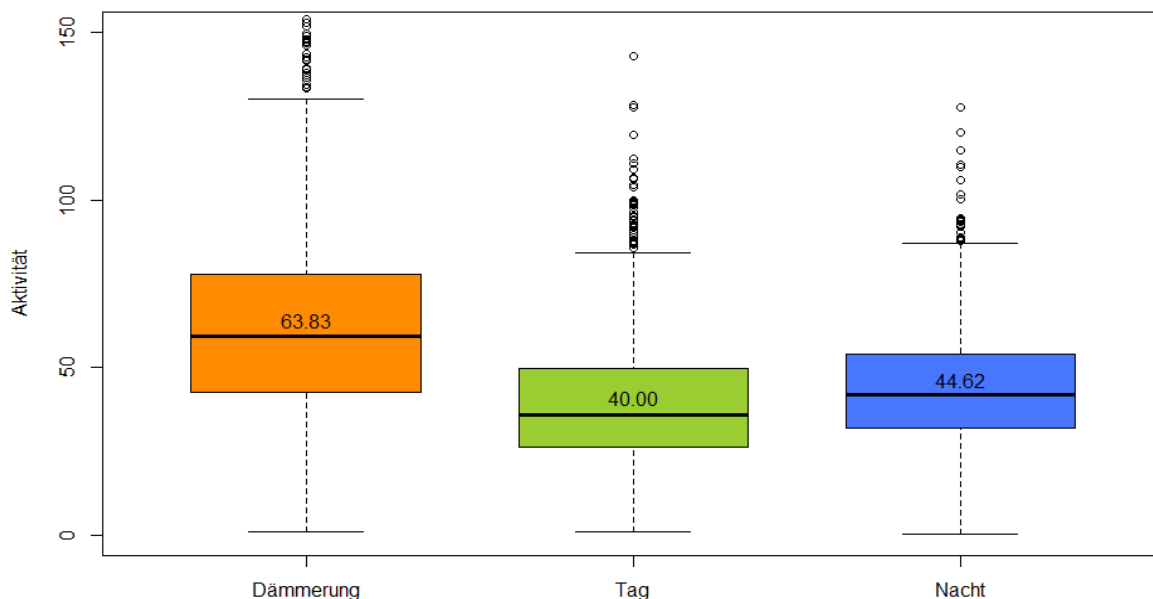


Abbildung 5: Boxplots der Aktivverteilung in den Tageszeiten

In der Grafik lässt sich erkennen, dass die Aktivität in der Dämmerung deutlich höher ist als in der Nacht und am Tage. Der Median der Aktivität in der Dämmerung liegt bei 63,83, bei der Nacht bei 44,62 und am Tag bei 40,00. Es gibt jedoch in allen drei Aktivitätsboxplots einige Ausreißer, was auf das individuelle Verhalten der einzelnen Rehe zurückzuführen ist. Der Korrelationskoeffizient nach Pearson zeigt zwischen den verschiedenen Aktivwerten folgende Ergebnisse:

Aktivität Dämmerung	:	Aktivität Nacht	=	0.4227
Aktivität Nacht	:	Aktivität Tag	=	0.5029
Aktivität Dämmerung	:	Aktivität Tag	=	0.8616

Daraus lässt sich schließen, dass ein starker, linearer Zusammenhang zwischen den Aktivitäten am Tag und in der Dämmerung besteht. Ein schwacher linearer Zusammenhang lässt sich zwischen den Aktivitäten in der Dämmerung und in der Nacht feststellen.

3.2 Aktivanteil

Der Aktivanteil gibt die Anteile von Aktiv und Ruhephasen als Prozentwert wieder. Wie in Abbildung 6 zu erkennen ist, ist auch der Anteil von Aktivität zur Ruhephase in der Dämmerung am höchsten mit einem Median von 72.28%, allerdings ist der Anteil der Aktivität am Tage höher als in der Nacht. Also genau das Gegenteil von dem Boxplots der Aktivitätswerte in Abbildung 5 aus Kapitel 3.1 „Aktivität“, dort sind die Aktivwerte der Nacht höher als die am Tage. Dies würde dafür sprechen, dass sich die Tiere in der Nacht weniger oft bewegen, aber dafür eine höhere Aktivität bei den Bewegungen aufweisen als am Tag.

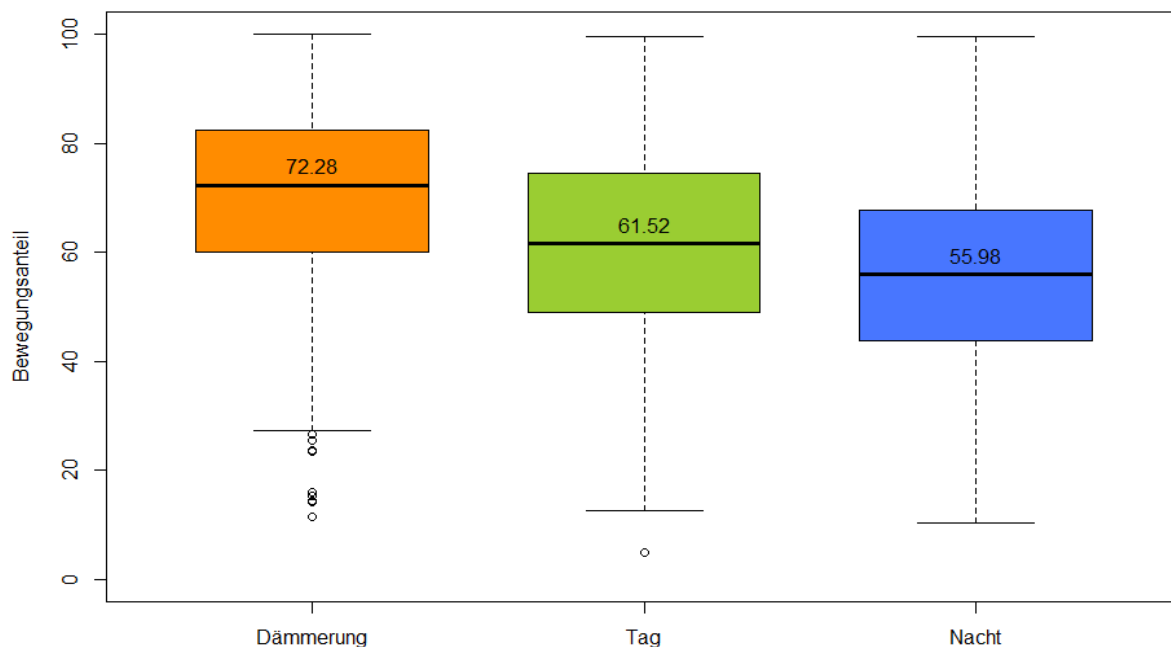


Abbildung 6: Boxplots der Bewegungsanteile in den Tageszeiten

Bei den Aktivanteilen in der Dämmerung gibt es einige Ausreißer, die einen sehr geringen Aktivanteil aufweisen. Zwischen den Anteilswerten der Aktivität in den Tageszeiten bestehen folgende Korrelationen:

Anteil Dämmerung	:	Anteil Nacht	=	0.7871
Anteil Nacht	:	Anteil Tag	=	0.7579
Anteil Dämmerung	:	Anteil Tag	=	0.4613

Die Bewegungsanteile der Dämmerung und der Nacht, sowie Tag und Nacht haben einen starken linearen Zusammenhang, während die Anteilswerte der Dämmerung und dem Tag einen mittleren linearen Zusammenhang aufweisen.

3.3 Luchsriss

Das Ziel dieser Bachelorarbeit ist es, mit dem Cox-Modell die Risikofaktoren der Rehe bezüglich eines Luchsrisses herauszuarbeiten. In diesem Kapitel sollen nun erst einmal die deskriptiven Zusammenhänge dargestellt werden und erste Schlussfolgerungen gezogen werden.

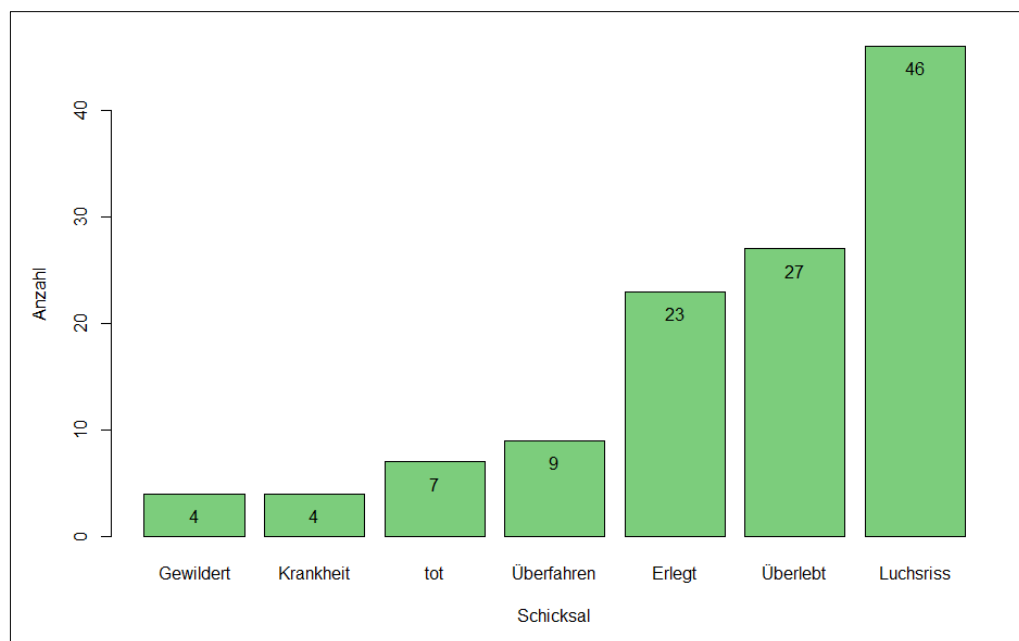


Abbildung 7: Schicksal der 120 untersuchten Rehe

In Abbildung 7 wurden die Schicksale der 120 untersuchten Rehe mit einem Balkendiagramm dargestellt. 46 der untersuchten Rehe wurden von einem Luchs gerissen, 27 Tiere haben zum Endzeitpunkt der Studie noch gelebt und wurden entsendert, 23 Tiere wurden erlegt und 9 überfahren. Bei 7 Tieren ist die Todesursache unbekannt, 4 Tiere sind einer Krankheit zum Opfer gefallen und weitere 4 Tiere sind von wildernden Hunden getötet worden. In der Studie sind insgesamt 93 Tiere gestorben, nur 27 der untersuchten Tiere haben überlebt. Von den verstorbenen Tieren sind 49,46%, also ungefähr die Hälfte, von einem Luchs gerissen worden. Insgesamt wurden 38% der beobachteten Rehe von einem Luchs gerissen.

3.3.1 Luchsriss und Alter

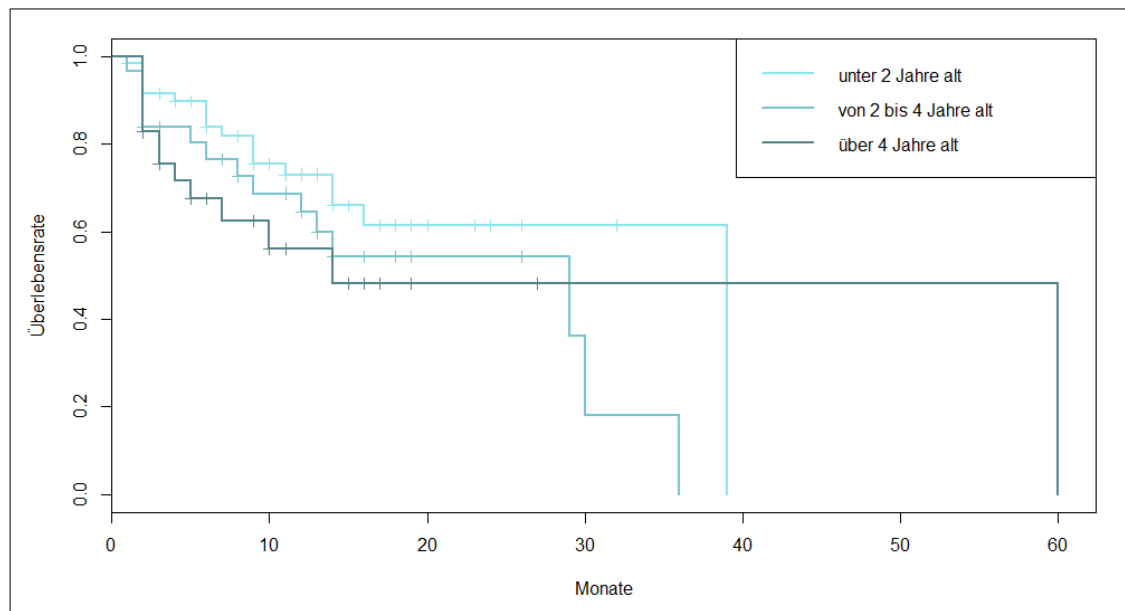


Abbildung 8: Kaplan-Meier Überlebenskurven des Alters bzgl. eines Luchsrisses

In Abbildung 8 wurden die Kaplan-Meier Überlebenskurven des Alters bzgl. eines Luchsrisses dargestellt. Dabei wurde das Alter zu Beginn der Studie in drei Intervalle eingeteilt: $(0,2]$, $(2,4]$ und $(4,10]$ Jahre. Am Anfang jeder Kurve haben die Tiere eine Überlebenswahrscheinlichkeit von 1, denn hier waren alle Tiere am Leben. Die „Treppenstufen“ bedeuten, dass ein Luchsriss stattgefunden hat, die kleinen Striche bedeuten, dass ein Tier „zensiert“ wurde. Ein zensiertes Tier starb entweder durch ein anderes Ereignis als ein Luchsriss, oder wurde entsendert. Die Kurve der unter zweijährigen Rehe ist oberhalb der anderen zwei Kurven, dies könnte ein Indiz dafür sein, dass diese Tiere ein höheres Survival haben. Die Tiere, die über 4 Jahre sind liegen zuerst unter den anderen zwei Kurven, am Ende gibt es jedoch einen „Ausreißer“. Bei dem Ausreißer handelt es sich um ein einziges Tier, das sehr lange beobachtet wurde und schließlich von einem Luchs gerissen worden ist. Da ab 30 Beobachtungsmonate nur noch sehr wenige Tiere unter Risiko standen, sind die Kurven ab dem 30. Monat mit Vorsicht zu betrachten. Insgesamt könnte man auf einen Alterseffekt *„je höher das Alter, desto höher das Risiko eines Luchsrisses“* schließen. Um Schlüsse aus dieser Grafik ziehen zu können, muss man jedoch bedenken, dass sie sehr stark von dem gewählten Intervalls Grenzen des Alters abhängt und nur als erster Überblick dient.

3.3.2 Luchsriss und Geschlecht

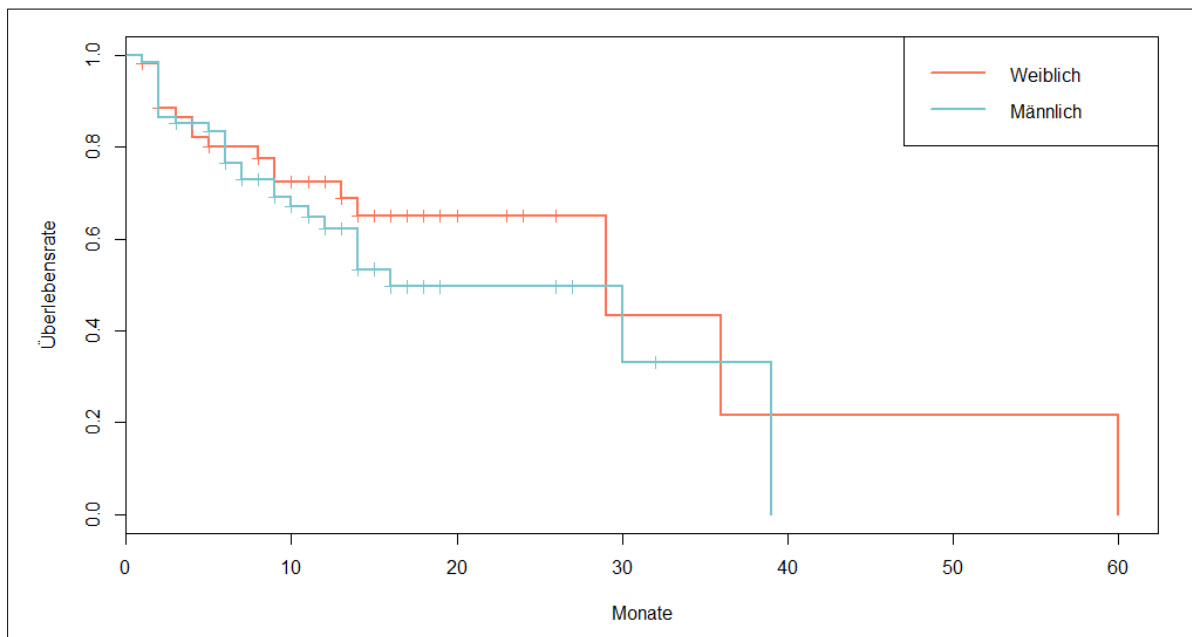


Abbildung 9: Kaplan-Meier Überlebenskurve des Geschlechts bzgl. eines Luchsrisses

In Abbildung 9 wurde ebenfalls zwei Kaplan-Meier Überlebenskurve geplottet. Dabei wurde das Geschlecht bzgl. des Luchsrisses betrachtet. Die rote Kurve ist die Überlebensrate der weiblichen Tiere und die blaue Kurve ist die Überlebensrate der männlichen Tiere. Zu Beginn der Studie liegen beide Kurven fast gleich auf, im Verlauf liegt dann die Kurve der weiblichen Tiere höher als die der männlichen Tiere. Gegen Ende der Untersuchung schneiden sich die beiden Kurven, dies liegt aber auch an den wenigen Tieren, die zu diesem Zeitpunkt noch unter Risiko standen. Wieder ist gut der Ausreißer zu erkennen. Da die beiden Kurven nah beieinander liegen und sich auch immer wieder schneiden, könnte man von keinem Effekt des Geschlechts auf die Überlebensrate bzgl. eines Luchsrisses ausgehen. Allerdings könnte auch eine Survivalanalyse des Cox-Modells mit sämtlichen Einflüssen aller Kovariablen zu einem anderen Ergebnis führen.

3.3.3 Aktivitätswerte und Luchsriss

Bereits im statistischem Praktikum zur Untersuchung der Aktivität bzgl. eines Luchsrisses wurde die Aktivität der Abenddämmerung analysiert (Zeis, et al., 2013). Im folgendem wird nun nicht nur die Aktivität der Abenddämmerung, sondern die Aktivität in Morgen- und Abenddämmerung, am Tag und in der Nacht bzgl. des Luchsrisses analysiert. An der folgenden Abbildung ist zu beachten, dass hier nicht ein einziger Aktivwert pro Reh geplottet wurde, sondern die Aktivitäten der einzelnen Beobachtungsintervalle der Rehe abgebildet wurden. Der Vorteil dieser Betrachtung wird im Kapitel 5.2 „Erweiterung des Cox-Modells mit zeitvariierenden Kovariablen“ noch genauer erläutert.

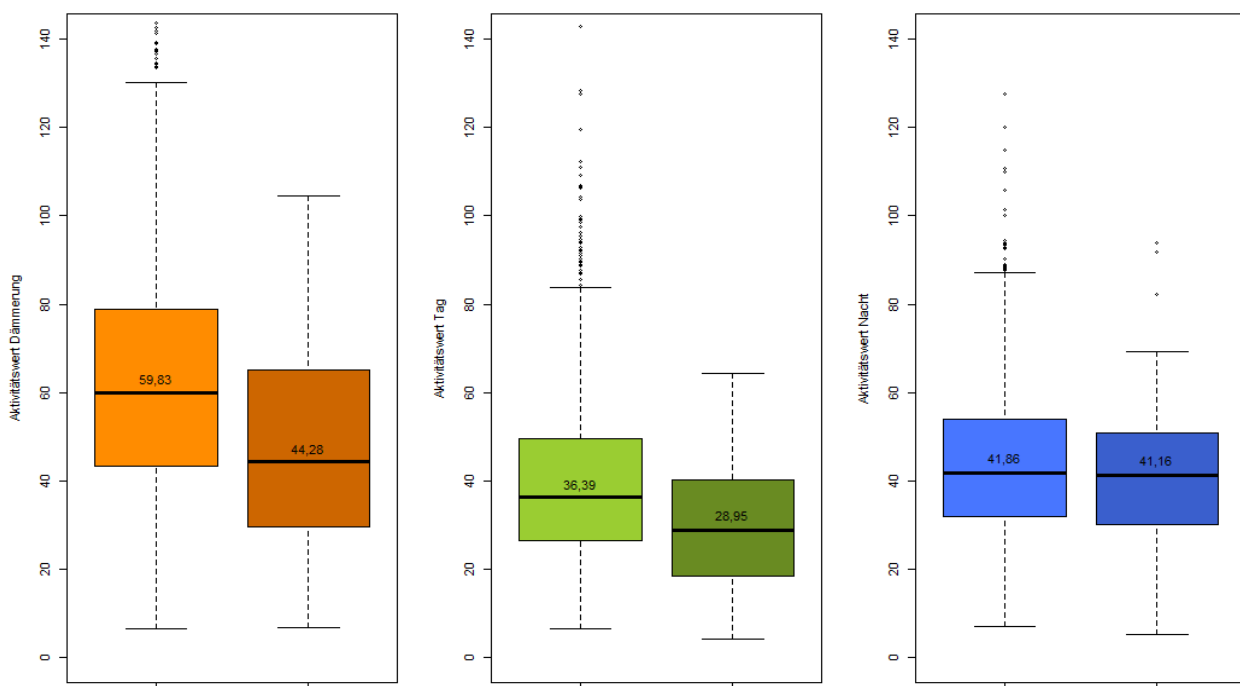


Abbildung 10: Darstellung der Aktivwerte bzgl. des Luchsrisses in den Tageszeiten

Wie bereits im Kapitel 3.1 „Aktivität“ deutlich zu erkennen war, liegt die Aktivität in der Dämmerung über der des Tages und der Nacht. Der erste Boxplot zeigt dabei jeweils die Aktivitätsverteilung der Rehe, die in dem beobachteten Intervall nicht vom Luchs gerissen worden sind und der zweite Boxplot die Aktivitätsverteilung der Rehe, die in dem beobachteten Intervall von dem Luchs gerissen worden sind. Bei der Aktivität in der Dämmerung ist ein starker Unterschied zwischen den Medianen bzgl. Luchsriss zu erkennen. Ein ähnliches, aber nicht so ausgeprägtes Bild bieten die beiden Boxplots der Tagesaktivität. Daraus könnte man schließen, dass die Tiere, die eine höhere Aktivität aufweisen ein geringeres Risiko haben von einem Luchs gerissen zu werden. Nur in der Nacht scheint es das keinen Unterschied bzgl. eines Luchsrisses und der Aktivität zu geben.

3.3.4 Aktivanteile und Luchsrisso

Während im Praktikum nur die Aktivwerte betrachtet wurden, werden in dieser Arbeit auch die Anteile der Bewegung zur Ruhephase untersucht. Auf die Kodierung und Berechnung der neuen Variablen wurde bereits im Kapitel 2.2.2 „*Bewegungsanteil*“ genauer eingegangen. An der folgenden Abbildung ist wieder zu beachten, dass es sich hierbei nicht um den Aktivanteil der einzelnen Rehe, sondern um den Aktivanteil der Rehe in den einzelnen Beobachtungsintervallen handelt.

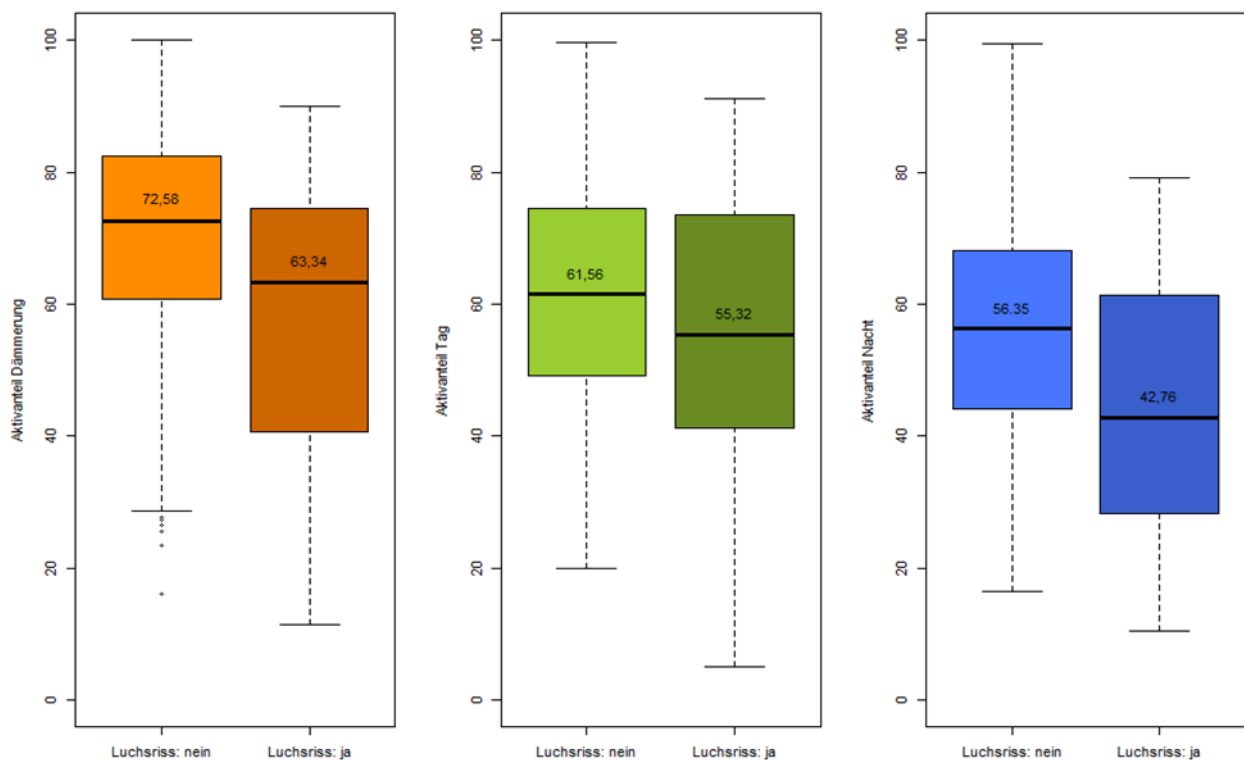


Abbildung 11: Darstellung der Aktivanteile bzgl. eines Luchsrisso in den Tageszeiten

In Abbildung 11 ist wiederum zu sehen, dass die Tiere mit einem höheren Aktivanteil eine geringere Wahrscheinlichkeit haben, vom Luchs gerissen zu werden. Allerdings zeigt sich hier, dass der Aktivanteil in der Nacht auf dasselbe Ergebnis kommt. Die Tiere, die in der Nacht einen höheren Aktivanteil haben, haben anscheinend eine deutlich geringere Wahrscheinlichkeit vom Luchs gerissen zu werden. Wie bereits in den vorherigen Kapiteln erwähnt unterscheiden sich die beiden Aktivitätsmessungen hier voneinander. Die Darstellungen der deskriptiven Analyse zeigen allerdings nur die einzelnen Einflüsse der Rehe bzgl. eines Luchsrisso. Ein viel deutlicheres Ergebnis liefert die Survivalanalyse unter Verwendung des Cox-Modells, in der sämtliche Einflüsse und Interaktionen einfließen.

4. Das Cox-Modell

Um den Einfluss von mehreren Kovariablen auf die Überlebenszeit zu untersuchen, wird in dieser Arbeit das Cox-Modell, auch bezeichnet als Proportionales Hazard Modell, verwendet. Als erstes soll aber in den folgenden Kapiteln die zu Grunde liegende Theorie erläutert werden.

4.1 Theorie zum Cox-Modell mit festen Kovariablen

Um Survivaldaten, wie die der vorliegenden Untersuchung der Rehe, untersuchen zu können benötigt es spezielle Techniken. Diese sind aus folgenden Gründen wichtig:

- Survivaldaten sind normalerweise nicht systematisch verteilt
- Nach Beendigung der Studie haben nicht alle Individuen ihren „Endpunkt“ erreicht. Der „Endpunkt“ ist das interessierende Ereignis, sei es der Tod des Individuums, die Heilung oder in dieser Untersuchung der Luchsriss. Daher wissen wir bei ihnen nicht die exakte Überlebenszeit. Diese Objekte werden als „zensiert“ bezeichnet. Im Fall der Rehe werden die Tiere, die nicht durch einen Luchsriss ums Leben gekommen sind, sondern aus anderen Gründen, ebenfalls als „zensiert“ betrachtet (Gesamter Abschnitt nach: (Everitt, et al., 2009)).

4.1.1 Die Survivalfunktion und Kaplan-Meier-Schätzer

Die Survivalfunktion gibt die Wahrscheinlichkeit an, dass die Überlebenszeit T größer oder gleich eines Zeitpunktes t ist. Also gibt sie die Wahrscheinlichkeit für ein Individuum an bis zu einem Zeitpunkt t überlebt zu haben (Lawless, 1982).

$$S(t) = P(T \geq t)$$

Gibt es Zensierungen im Datensatz wird die Survivalfunktion normalerweise durch den Kaplan-Meier-Schätzer geschätzt. Voraussetzung für den Kaplan-Meier-Schätzer ist das vorherige Ordnen der Überlebenszeiten: $t_1 \leq t_2 \leq \dots \leq t_n$.

$$\hat{S}(t) = \prod_{j:t_{(j)} \leq t} \left(1 - \frac{d_j}{r_j}\right)$$

In der Formel beschreibt r_j die Anzahl der Individuen, die unter Risiko stehen, vor dem Zeitpunkt t_j einschließlich der zensierten Individuen zu diesem Zeitpunkt. d_j ist die Anzahl der Individuen, bei denen zum Zeitpunkt t_j das Ereignis eingetreten ist.

Bei einem Zeitpunkt $t = 0$ wäre der Kaplan-Meier-Schätzer $\hat{S}(0) = 1$, da zum Zeitpunkt 0, den Anfang der Untersuchung, alle Individuen überlebt haben.

Die Varianz des Kaplan-Meier-Schätzers ist definiert durch:

$$Var(\hat{S}(t)) = (\hat{S}(t))^2 \sum_{j:t_{(j)} \leq t} \frac{d_j}{r_j(r_j - d_j)}$$

Gesamter Abschnitt (Everitt, et al., 2009)

4.1.2 Die Hazard-Funktion

Die Grundlage für die Untersuchung mit dem Cox-Modell ist die Hazard-Funktion. Diese wird auch als Hazard-Rate bezeichnet und ist definiert durch:

$$h(x) = \lim_{s \rightarrow 0} \frac{P(t \leq T < t + s \mid T \geq t)}{s}$$

Man kann also $h(x)$ als die approximative Wahrscheinlichkeit ansehen, dass bei dem Individuum zum Zeitpunkt x das Ereignis eintritt unter der Bedingung, dass das Individuum zu Beginn des Intervalls überlebt hat. Es gibt viele allgemeine Formen der Hazard-Rate, jedoch mit der Einschränkung, dass $h(x)$ keinen negativen Wert annehmen darf: $h(x) \geq 0$

Gesamter Abschnitt (Klein, et al., 1997)

4.1.3 Das Cox-Modell mit festen Kovariablen

Das Cox-Modell wird auch als „Proportionales Hazard Modell“ bezeichnet und wurde 1972 von Sir David Cox entwickelt. Dieses Modell ist das meistbenutzte Modell, wenn es darum geht eine Beziehung von mehreren Kovariablen auf die Überlebenswahrscheinlichkeit und andere zensierte Ergebnisse (z.B. eine Heilung von einer Krankheit) zu berechnen

Sei X_{ij} , die j -te Kovariable ($j = 1, \dots, p$) des i -ten Individuums ($i = 1, \dots, n$). Aus einem Datensatz mit n Individuen und ihren p Kovariablen entsteht so eine $n \times p$ Matrix. In dieser Matrix steht somit Zeile X_i für den Kovariablenvektor des i -ten Individuums. Wir nehmen nun an, dass jede Kovariable von der Zeit unabhängig ist und bezeichnen diese als „feste Kovariable“. Als Beispiel für eine zeitunabhängige Kovariable ist das Geschlecht oder Geburtsgewicht zu nennen.

Das Cox-Modell geht davon aus, dass sich die Kovariablen auf die Hazard-Funktion des Individuums wie folgt auswirken:

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_1 x_1 + \dots + \beta_p x_p)$$

Dabei handelt es sich bei $\lambda_0(t)$ um eine unspezifische, nicht negative Funktion der Zeit, die auch als *baseline hazard* bezeichnet wird. Im hinteren Term sind nun die Effekte der Kovariablen enthalten. β ist ein $p \times 1$ -Spalten Vektor der Koeffizienten. Der Schätzer für β basiert auf der *partial likelihood function*, die ebenfalls von Sir David Cox entwickelt wurde (Gesamter Abschnitt nach (Therneau, et al., 2001)).

Die Kovariablen haben dabei einen loglinearen Einfluss, dieser Einfluss wirkt sich dann multiplikativ auf die Hazardrate aus (Potuschak, 2000).

Partieller Likelihood für den Parameter β :

Das Individuum mit dem Kovariablenvektor x_i erlebt bei t_i das interessierende Ereignis (Tod, Heilung usw.) Die „event“-Zeiten $t_1 < \dots < t_i < \dots < t_k$, also die Zeitdauer der Individuen bei denen das interessierende Ereignis eingetreten ist ($k \leq n$), werden geordnet. Die Individuen, die noch vor t beobachtet (also auch die Individuen die noch nicht zensiert sind) wurden, sind in der Risikomenge R_t . Daraus ergibt sich PL_β :

$$PL_\beta = \prod_{i=1}^k \frac{\exp(x_i \beta)}{\sum_{j \in R(t(i))} \exp(x_j \beta)}$$

Und somit der Maximum PL-Schätzer:

$$\hat{\beta} = \arg \max_{\beta} \log PL(\beta)$$

(Gesamter Abschnitt (Fahrmeir, 2007))

4.1.4 Hazard Ratio

Die *hazard ratio* dient nun dazu zwei Gruppen oder zwei Individuen zu vergleichen. Sie setzt zwei Hazard Raten in Beziehung zueinander. Wenn alle anderen Kovariablen festgehalten werden kann folgender Zusammenhang, der über die Zeit hinweg konstant ist, festgestellt werden:

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\cancel{\lambda_0(t)} \exp(\beta x_i)}{\cancel{\lambda_0(t)} \exp(\beta x_j)} = \frac{\exp(\beta x_i)}{\exp(\beta x_j)}$$

Ein *hazard ratio* von 1 bedeutet, dass das Risiko für Individuum eins und für Individuum zwei gleich groß ist, ein *hazard ratio* von größer 1 bedeutet ein höheres Risiko für Individuum eins und ein *hazard ratio* von kleiner 1 bedeutet ein geringeres Risiko für Individuum eins als Individuum zwei. Ist nun also das Risiko des Ereignisses (sei es Tod oder Heilung) zu einem bestimmten Zeitpunkt zweimal so hoch für Individuum eins als für Individuum zwei, so bleibt dieses Risiko auch zweimal so hoch über alle anderen Zeitpunkte. Daher der Name des „*Cox's proportional hazards model*“ (Everitt, et al., 2009).

4.1.5 Bindungen

Der PL-Schätzer wurde entwickelt unter der Annahme von kontinuierlichen Daten, allerdings trifft dies nicht auf wahre Datenstrukturen zu. Die Daten beinhalten oft Bindungen zwischen den interessierenden Ereignissen. Wenn ein Individuum im Beobachtungszeitraum (4, 5] stirbt (das interessierende Ereignis wäre in diesem Beispiel der Tod) und andere Individuen in demselben Zeitraum gestorben sind, spricht man von einer Bindung. Es existieren vier Varianten, die dieses Problem bei der Berechnung solcher verbundenen Daten lösen (Therneau, et al., 2001):

- Breslow approximation
- Efron approximation
- Exact partial likelihood
- Averaged likelihood

Die Cox-Modelle in dieser Arbeit werden mit der *Efron approximation* berechnet, da sie als exakter gilt und in dem Statistikprogramm R unter der Verwendung des Package „Survival“ als *default* angegeben ist:

Dabei sind $t_1 < \dots < t_i < \dots < t_k$ die geordneten Zeiten, in denen ein Ereignis stattgefunden hat, d_i ist die Anzahl der interessierenden Ereignisse zum Zeitpunkt t_i und D_i ist die Menge der Individuen, die das interessierende Ereignis zum Zeitpunkt t_i haben.

$$PL_{Efron}(\beta) = \prod_{i=1}^k \frac{\exp(s_i \beta)}{\prod_{j=1}^{d_i} \left[\sum_{k \in R(t_i)} \exp(x_k \beta) - \frac{j-1}{d_i} \sum_{k \in D_i} \exp(x_k \beta) \right]}$$

$$\text{wobei: } s_i = \sum_{j \in D_i} x_j$$

Die *Efron approximation* bezieht also die verbundenen Ereignisse mit Gewichtungen in den Schätzer ein. Wenn keine Bindungen zwischen den interessierenden Ereignissen bestehen, dann: $PL_{Efron}(\beta) = PL_{\beta}$. (Höhle, 2009)

4.2 Theorie zum Cox-Modell mit zeitvariierenden Kovariablen

Im Kapitel 4.1.3 wurde bereits die Theorie zum Cox-Modell mit festen Kovariablen erläutert. Dieses Cox-Modell soll nun mit zeitvariierenden Kovariablen erweitert werden. Die Verwendung von zeitvariierenden Kovariablen bietet sich vor allem in Studien an, in denen zu mehreren Zeitpunkten Messungen durchgeführt wurden, wie z.B. Labortests. Der Grundgedanke ist hierbei, dass man die Daten erweitert, um eine genauere Untersuchung der Einflüsse machen zu können.

4.2.1 Zählprozess und Zeitskalen

Essentiell für die Erweiterung des Cox-Modells ist der Zählprozess. Dabei geht man nun nicht mehr davon aus, dass eine Zensierung eine unvollständige Messung ohne Endpunkt ist, sondern, dass das Ereignis bis jetzt noch nicht eingetreten ist, der „Geiger-Zähler“ aber weiterhin „tickt“. Für das Cox-Modell gibt es verschiedenen Arten von Zeitskalen, die für die Zählprozesse verwendet werden können:

• *Zeit in der Studie*

• *Zeit als Alter*

• *Zeit seit Diagnose*

• *Kalenderzeit*

Die Zeitskala sollte zu den Daten passend gewählt werden. Wenn z.B. eine Krankheit untersucht werden soll und das interessierende Ereignis der Tod durch die Krankheit ist, bietet sich natürlich an, die Zeit seit Diagnose oder erste Symptome der Krankheit als Skala zu verwenden. Die Interpretation von Baseline-Hazard hängt von der Zeitskala ab

Das Vorgehen des Zählprozesses bewirkt, dass nun nicht mehr eine Zeile pro Individuum im Datensatz existiert, sondern eine Reihe von Beobachtungen. Jede dieser Beobachtungen findet in einem *linksoffenen* und *rechtsgeschlossenen* Zeitintervall (start, stop] statt. Dies ist wichtig, um ein Überlappen zu verhindern. Außerdem werden dadurch wichtige, mathematische Einschränkungen beachtet: Wenn zum Beispiel zwei Intervalle von (0,15] (15,30] gehen und eine Berechnung zum Zeitpunkt $t=15$ gemacht werden soll, werden nur die Kovariablendaten genutzt, die vor dem Zeitpunkt $t=15$ gemessen worden sind. Somit werden die Daten und die „unter Risiko-Intervalle“ zu vorhersagbaren Prozessen. Aus diesem Grund ist auch zu beachten, dass keine zukünftigen Ereignisse in einer Kovariable auftauchen.

Der Einwand, dass die einzelnen Beobachtungen eines Individuums korrelieren und somit nicht verwendet werden dürfen, lässt sich entkräften durch die Tatsache, dass das Cox-Modell diese Problem mit internen Berechnungen löst: Für jedes einzelne interessierende Ereignis (Tod, Heilung usw.) werden die Fälle, die im selben Beobachtungsintervall „unter Risiko“ stehen, aufsummiert. Da sich die Beobachtungsintervalle niemals überlappen ist gewährleistet, dass jedes Individuum höchstens einmal in der Berechnung für dieses Zeitintervall vorkommt. Die Beobachtungen der einzelnen internen Berechnungen sind somit unabhängig voneinander. Die verwendeten Variablen „start“ und „stop“ sind sozusagen ein Trick, damit die Beobachtungszeiträume für jedes Individuum korrekt zugeordnet werden kann.

4.2.2 Das Cox-Modell mit zeitvariierenden Kovariablen

Bei den zeitvariierenden Kovariablen handelt es sich, wie der Name bereits erklärt, um Variablen, die sich mit der Zeit verändern. Ein Beispiel für eine zeitvariierende Kovariable wären Blutdruck, Cholesterin, BMI usw. Diese Kovariablen stehen nicht von Anfang an fest, sondern können sich im Laufe der Zeit verändern (Klein, et al., 1997).

Das Cox-Modell mit zeitvariierenden Kovariablen lässt sich nun mit folgender Formel darstellen:

$$\lambda_i(t) = \lambda_0(t) \exp(\beta_1 x_1(t) + \dots + \beta_p x_p(t))$$

Dabei gibt es im Gegensatz zum Cox-Modell mit festen Kovariablen die Struktur mit der zeitvariierenden Komponente $x_i(t)$ erweitert. In das Modell können sowohl zeitvariierende als auch feste Kovariablen aufgenommen werden. Bei den festen Kovariablen bleibt dann natürlich die Ausprägung für jedes Zeitintervall gleich (Klein, et al., 1997).

Die Berechnung des Partiellen Likelihood Schätzer geht man fast analog wie zum Modell mit festen Kovariablen vor, allerdings wird hier der Kovariablenvektor zum Zeitintervall (t_i) betrachtet:

$$PL_{\beta} = \prod_{i=1}^k \frac{\exp((x_i(t_i))\beta)}{\sum_{j \in R(t_i)} \exp((x_j(t_i))\beta)}$$

Allerdings muss, um die Gültigkeit von PL_{β} zu gewährleisten, die Voraussetzung erfüllt sein, dass $x(t_i)$ ein vorhersehbarer, stochastischer Prozess ist. Wenn eine Variable nicht nur variierend ist, sondern zeitabhängig, ist diese Annahme verletzt und der Schätzer verliert seine Gültigkeit. Ein Test, der eine PH-Annahmen Verletzung aufdeckt, ist der Grambsch und Therneau Test. Um eine Variable, die zeitabhängig ist, ins Modell aufnehmen zu können, muss folgende Modellierung vorgenommen werden:

Aus $x_1 \beta_1$ wird $x_1 \cdot (\beta_1 + \beta^* t) = x_1 \beta_1 + t x_1 \cdot \beta^*$, wobei $t x_1$ die neue, zeitabhängige Kovariable ist. Die beiden Parameter β_1 und β^* können dann mit dem PL_{β} -Ansatz geschätzt werden.

Absatz nach (Fahrmeir, 2007)

4.3 Residuen und Test auf PH-Annahme

Im Cox-Modell finden vor allem die Cox-Snell-Residuen, die Martingalen-Residuen, die Schoenfeld-Residuen und die Devianz-Residuen Verwendung. Die Cox-Snell-Residuen dienen dazu, die Eignung des Modells zu überprüfen. Durch die Marginalresiduen wird die funktionale Form der Variablen überprüft und die Devianz-Residuen sind eine Hilfe bei der Erkennung von Ausreißern. In diesem Kapitel sollen nun kurz die Residuen erläutert werden, die für die Datenanalyse der Rehe verwendet werden, außerdem soll in diesem Kapitel kurz auf die Theorie des Grambsch und Therneau Test eingegangen werden, dieser findet Verwendung beim Testen auf der PH-Annahme.

4.3.1 Cox-Snell-Residuen

Für Individuen $i = 1, \dots, n$ mit Survivalzeit T_i , Kovariablenvektor \mathbf{x}_i und kum. Hazardrate $\Lambda(T_i|\mathbf{x}_i)$ ist

$$Y_i = \Lambda(T_i|\mathbf{x}_i) \sim \text{Exp}(1)$$

Im Cox-Modell gilt:

$$\Lambda(t_i|\mathbf{x}_i) = \Lambda_0(t_i)\exp(\mathbf{x}_i\beta)$$

Ersetze nun β durch PL-Schätzer $\hat{\beta}$ und $\Lambda_0(t)$ durch Breslow-Schätzer $\hat{\Lambda}_0(t)$. Damit sind die Cox-Snell-Residuen:

$$r_i = \hat{\Lambda}_0(t) \exp(\mathbf{x}_i\hat{\beta}), i = 1, \dots, n$$

r_i sollte $\approx \text{Exp}(1)$ verteilt sein. Wende Nelson-Aalen-Schätzer auf „Pseudo-Lebensdauern“ $r_i, i=1, \dots, n$ an. Wenn der Nelson-Aalen-Schätzer deutlich von t abweicht ist dies ein Hinweis auf Modellverletzung

(Kapitel nach (Fahrmeir, 2007)).

4.3.2 Martingale Residuen

Sei $Y_i(t)$ der Indikator dafür, dass ein Individuum unter Risiko zum Beobachtungszeitpunkt t steht, $N_i(t)$ ist die kumulative Anzahl von Ereignissen des Individuums bis zum Zeitpunkt t , $\bar{N}(t) = \sum N_i$ die gesamte Anzahl der Ereignisse und $r_i(t) = \exp(\mathbf{x}_i(t)\beta)$.

Wenn $\Lambda_i(t)$ die wahre kumulative Ereignisrate für Individuum i ist, dann ist der erwartete, beobachtete Prozess

$$M_i(t) = \int Y_i(t) \{dN_i(t) - d\Lambda_i(t)\}$$

das Individuum-spezifische Martingal. Nach Herleitung (siehe (Therneau, 1996)) sind somit die Martingale-Residuen zum Zeitpunkt t :

$$M_i(t) = N_i(t) - \int_0^t r_i(\beta, s) Y_i(s) d\hat{\Lambda}_0(\beta, s).$$

Martingal-Residuen sind dabei eine einfache Differenz im Intervall $(0, t]$ zwischen der beobachteten Anzahl von Ereignissen für ein Individuum und der erwarteten Anzahl durch das Modell.

Gesamtes Kapitel nach (Therneau, 1996)

4.3.3 Grambsch und Therneau Test

Die Hauptannahme des Cox-Modells ist der proportionale Hazard. Im Modell mit festen Kovariablen ist dies die Hazard Ratio-Annahme, die unabhängig von der Zeit ist. Bei zeit-variierenden Kovariablen ist der Hazard Ratio für zwei Individuen:

$$\frac{\exp(\mathbf{x}_i(t)\beta)}{\exp(\mathbf{x}_j(t)\beta)}$$

Diese Annahme ist jedoch nicht unabhängig von der Zeit. Allerdings ist die relative Auswirkung von zwei Werten einer Kovariable in einem Koeffizienten β zusammengefasst.

Um zu testen, ob eine Verletzung der PH-Annahme vorliegt, wird der Grambsch und Therneau Test verwendet. Dabei werden künstlich erzeugte zeitabhängige Koeffizienten erzeugt und diese auf die Variablen getestet. Bei einem $p \leq 0,05$ liegt eine Verletzung der PH-Annahme vor und die Variable muss evtl. geschichtet werden oder ein anderes Modell gerechnet werden. Genaue Herleitung und Annahmen des Tests siehe „*Modeling Survival Data*“ von Terry Therneau und Patricia Grambsch, (Therneau, et al., 2001))

Eine Schichtung sollte vorgenommen werden, falls die PH-Annahme verletzt ist oder inhaltliche Gründe dafür sprechen. Eine Schichtung ist nur bei Variablen möglich, die kategoriale oder diskret metrische Ausprägungen besitzen. Dabei wird für jede Schicht eine Hazardrate berechnet.

$$\lambda_s(t) = \lambda_{0_s}(t) \exp(\beta_1 x_1(t) + \dots + \beta_p x_p(t))$$

Eine Schichtung von variierenden Kovariablen ist nicht möglich. Durch die Schichtung verändert sich die Behandlung von zeitvariierenden und zeitabhängigen Kovariablen nicht, auch die zensierten oder verbundenen Werte behalten ihre Behandlung bei. Die Bindungen können sogar aufgelöst werden (Potuschak, 2000).

5. Datenanalyse des Rehdatensatzes

Im Kapitel 5 werden nach vorangegangenen Untersuchungen zur Korrelation zwischen den Kovariablen die Modelle gefittet. Dabei wird zuerst das Modell mit den festen Kovariablen berechnet und anschließend werden die Modelle mit variierenden Kovariablen berechnet. Am Ende des Kapitels werden die Modelle auf Eignung und auf Verletzung der Proportional-Hazard-Annahme (PH-Annahme) mittels dem Grambsch und Therneau Test und den Residuen geprüft.

5.1 Korrelationen zwischen den Kovariablen

Das Kodieren der Faktoren und ihre Interaktionen untereinander folgen den gewöhnlichen Regeln des Linearen Modells. Falls ein starker Zusammenhang zwischen zwei Variablen besteht, sollten diese entweder in einer logischen Weise zusammengefügt oder gleich zwei getrennte Modelle gerechnet werden. Bei zusammengeführten Variablen ist besondere Aufmerksamkeit beim Interpretieren dieses Einflusses geboten. (Klein, et al., 1997)

In dem Cox-Modell mit festen Kovariablen stehen folgenden Variablen zur Verfügung

Kovariable	Einheiten
Geschlecht	m / w
Alter	Jungtier / Erwachsen
Aktivwert Dämmerung	13,09 – 144
Aktivwert Tag	10,48 – 80,04
Aktivwert Nacht	18,17 – 89,67
WSI.Geburtswinter	0 - 120

Tabelle 5: Kovariablen des Cox- Modells

Im erweiterten Cox-Modell stehen die folgenden Variablen zu Verfügung:

Kovariable	Einheiten	variierend / fest	Änderung
Geschlecht	m / w	fest	Nie
Alter	6 – 91 Monate	fest	Nie
Aktivwert Dämmerung	6.50 – 203.80	variierend	Monatlich
Aktivwert Tag	4.09 – 142.80	variierend	Monatlich
Aktivwert Nacht	5.02 – 127.60	variierend	Monatlich
Aktivanteil Dämmerung	11.41% – 100%	variierend	Monatlich
Aktivanteil Tag	4.99% – 99.56%	variierend	Monatlich
Aktivanteil Nacht	10.44% – 99.46%	variierend	Monatlich
WSI	0 - 34	variierend	Monatlich*

*in den Wintermonaten

Tabelle 6: Kovariablen des erweiterten Cox-Modells

Wie im Kapitel zur Theorie des Cox-Modells bereits erwähnt, dürfen zwei Variablen, die stark miteinander korrelieren, nicht in ein Modell aufgenommen werden. Dieses Problem wird als „Multicollinearity“ bezeichnet. Die am weitesten verbreitete Diagnostik für diese Probleme ist der „*Varianz Inflations Faktor*“ kurz VIF genannt. Dabei wird für jede Variable ein lineares Modell gegen die anderen Variablen gerechnet und das Bestimmtheitsmaß R^2 für diese Variable ausgegeben. Die VIF lässt sich nun folgendermaßen berechnen:

$$VIF_j = \frac{1}{(1 - R_j^2)}$$

Wobei R_j^2 das Bestimmtheitsmaß des linearen Modells der j-ten Variable auf die restlichen Variablen ist. Wenn nun eine Variable einen VIF von 1,6 hätte, würde das bedeuten, dass die Varianz des Koeffizienten 60% höher wäre, als wenn die Variable mit den anderen Einflussgrößen unkorreliert wäre. Der VIF hat eine untere Grenze von 1, ein Wert nahe der 1 bedeutet also, dass die Variable fast vollständig unkorreliert mit den anderen Variablen ist. Über den Grenzwert, wann eine Variable nicht mehr mit den anderen Variablen in das Modell aufgenommen werden sollte, gibt es bei den Experten Unstimmigkeiten. In dieser Bachelorarbeit möchte ich jedoch Prof. Paul Allison folgen, der einen Wert von 2.50 vorschlägt, dieser Wert würde ein Bestimmtheitsmaß von 0,6 entsprechen (Allison, 2012).

Die Varianz-Inflations-Faktoren für die **festen Kovariablen** im Cox-Modell liefert folgende Tabelle:

	Aktivwert Dämmerung	Aktivwert Tag	Aktivwert Nacht	WSI
VIF (gesamt)	7,02	7,76	1,38	1,04
VIF (Mod.1f)	1,23	-	1,27	1,03
VIF (Mod. 2f)	-	1,36	1,43	1,04

Tabelle 7: Varianz-Inflations-Faktor für Gesamtmodell, Modell 1f und Modell 2f mit festen Kovariablen

Der VIF-Wert der Kovariable „Aktivwert Dämmerung“ beträgt 7,02 er ist somit > als 2,5, auch der VIF-Wert der Kovariable „Aktivwert Tag“ hat einen zu hohen Wert von 7,76. Im Modell mit festen Kovariablen liegt ein multikollinearitäts Problem vor, deshalb werden zwei getrennte Modelle 1f und 2f gerechnet. Wie man in der Tabelle sehen kann, liegt nach der Trennung der Modelle kein multikollinearitäts Problem vor, die Variablen dürfen verwendet werden. Die Bezeichnung 1f steht für erstes Modell mit festen Kovariablen.

Im erweiterten Cox-Modell werden die Aktivwerte und die Aktivanteilswerte aus inhaltlichen Gründen nie in demselben Modell berechnet, deshalb werden von vornherein auch zwei VIF-Tabellen erstellt.

Die Varianz-Inflations-Faktoren für die **Aktivwerte** liefert folgende Tabelle:

	Aktivwert Dämmerung	Aktivwert Tag	Aktivwert Nacht	Alter	Habitat	Geschlecht	WSI
VIF (gesamt)	4,28	4,65	1,55	1,04	1,31	1,21	1,11
VIF (Mod. 1)	1,60	-	1,30	1,04	1,26	1,17	1,11
VIF (Mod. 2)	-	1,73	1,53	1,04	1,26	1,11	1,07

Tabelle 8: Varianz-Inflations-Faktor für Gesamtmodell, Modell 1 und Modell 2 mit Aktivwerten

Wie zu sehen ist, dürfen die zwei Aktivwerte Dämmerung und Tag nicht in dasselbe Modell, die VIF für diese Variablen sind $> 2,5$. Dadurch werden zwei verschiedene Modelle gerechnet. In diesen beiden Modellen werden nochmals nach den gleichen Regeln die VIF der Variablen bestimmt. Aus der Tabelle 7 geht schließlich hervor, dass in Modell 1 und in Modell 2 kein multikollinearitäts Problem besteht.

Die Varianz-Inflations-Faktoren für die **Aktivanteile** liefert folgende Tabelle:

	Aktivanteil Dämmerung	Aktivanteil Tag	Aktivanteil Nacht	Alter	Habitat	Geschlecht	WSI
VIF (gesamt)	3,87	3,32	7,22	1,08	1,22	1,12	1,12
VIF (Mod. 1a)	1,24	-	-	1,03	1,16	1,12	1,11
VIF (Mod. 2a)	-	1,07	-	1,04	1,09	1,09	1,03
VIF (Mod. 3a)	-	-	1,15	1,03	1,16	1,09	1,06

Tabelle 9: Varianz-Inflations-Faktor für Gesamtmodell, Modell 1a und Modell 2a mit Aktivanteilen

Auch bei dem Modell mit den Aktivanteilen würde es ein multikollinearitäts Problem geben. Der VIF der Variable des Aktivanteils in der Nacht liegt sogar bei 5,69. Dadurch werden drei getrennte Modelle gerechnet. Damit es zwischen den Modellen nicht zu Verwechslungen kommt, werden ab sofort alle Modelle mit Aktivanteilen mit 1a, 2a und 3a benannt.

Die Korrelationen der kategorialen Merkmale wurden bereits im Vorfeld getestet, dabei konnte kein starker linearer Zusammenhang mit den anderen Variablen festgestellt werden.

Um das Cox-Modell berechnen zu können wird die Tabelle des Rehdatsatzes mit dem Statistikprogramm R unter Verwendung des package „*Survival*“ gefittet. Nachdem alle Kovariablen, bei denen kein multikollinearitäts Problem festgestellt wurde, in ein Modell aufgenommen worden sind, wird mit der R-Funktion „*step*“, das **AIC** – „*Akaike's Informationskriterium*“ berechnet. Dabei wird aus dem Modell mit vorwärts und rückwärts Selektion (*method* = „*both*“) die Variablen entfernt, die keinen signifikanten Einfluss haben. Die Variablenselektion sucht dabei das Modell mit dem größten Vorhersagewert, deshalb können auch Variablen im Modell verbleiben, die auf einem 10% Niveau signifikant sind. Je kleiner der AIC, desto besser ist das Modell. Somit findet auf diesen Weg eine Variablenselektion statt.

Der AIC für das Cox-Modell lässt sich mit folgender Formel berechnen:

$$AIC = -2 \log PL(\hat{\beta}) + 2p$$

Wobei $p = \dim(\beta)$ und die Komplexität des Modells bestraft (Fahrmeir, 2007).

5.2 Cox-Modell mit festen Kovariablen

Wie bereits im vorangegangenen Kapitel gezeigt, werden im Falle von ausschließlich festen Kovariablen zwei getrennte Modelle gerechnet. Da in dieser Arbeit das Augenmerk auf die Modelle mit zeitvariierenden Kovariablen gelegt ist, wird in diesem Kapitel nur auf das Modell 1f eingegangen, Modell 2f befindet sich im Anhang. Auch wurde die Variablen Habitat und Aktivanteil ausschließlich in den Modellen mit zeitvariierenden Kovariablen berechnet.

5.2.1 Modell 1f

Im statistischen Vorpraktikum wurde bereits das Cox-Modell mit festen Kovariablen gerechnet. Da es sich aber um feste Kovariablen handelt, konnte der Aktivwert in der Dämmerung nur einmal pro Reh bestimmt werden, das Alter nur zum Zeitpunkt der Besenderung in Unterteilung Jungtier - erwachsenes Tier verwendet werden und der WSI des ersten Winters aufgenommen werden (Zeis, et al., 2013).

Call:

```
coxph (formula = Surv(start, stop, event) ~ Alter + Geschlecht  
+ Aktiv_daem + Aktiv_nacht + WSI + Alter_Besenderung, data = Tabelle)
```

Nach der Variablenselektion mittels der Funktion „step“ wird folgende Tabelle ausgegeben:

	Coef	exp(coef)	Pr(> z)
Aktiv Dämmerung	-0.0568	0.9448	p < 0,0001
Aktiv Nacht	0.0266	1.0269	0.0411
Geschlecht männl.	1.0429	2.8375	0.0033
WSI Geburtswinter	-0.0102	0.9898	0.0166
Alter Jungtier	-0.6289	0.5332	0.0481

Tabelle 10: Einflüsse der festen Kovariablen im Modell

Nach der Variablenselektion mit der Funktion „step“ wurde keine Variable aus dem Modell entfernt.

Zur Visualisierung wurden die Einflüsse der Kovariablen mittels Grafikplots dargestellt.

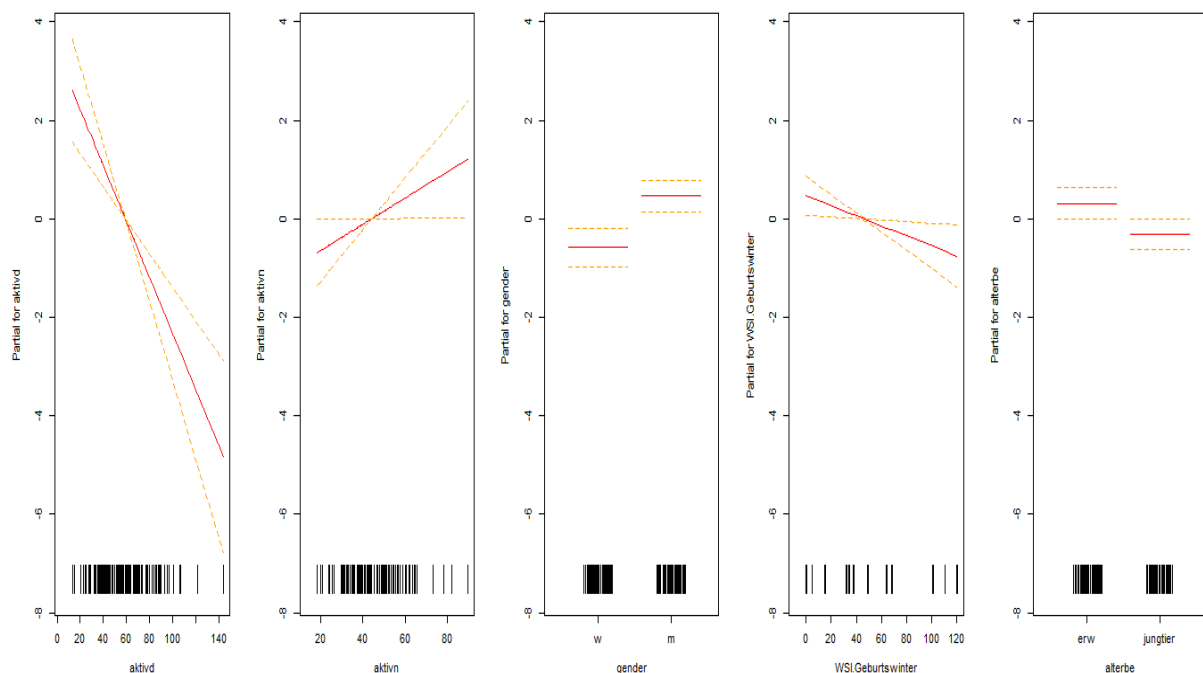


Tabelle 11: Partieller Einfluss der Kovariablen im Modell 1f

Interpretation:

- Wenn die Aktivität in der Dämmerung um **zehn** Einheiten ansteigt, so sinkt die Wahrscheinlichkeit eines Luchsrisses um den Faktor $\exp(-0,568) = 0,57$.
- Wenn die Aktivität in der Nacht um zehn Einheiten ansteigt, so steigt die Wahrscheinlichkeit eines Luchsrisses um den Faktor $\exp(0,266) = 1,30$.
- Im Gegensatz zu einem weiblichen Tier, erhöht sich die Wahrscheinlichkeit eines Luchsrisses für ein männliches Tier um den Faktor $\exp(1.0429) = 2,84$.
- Erhöht sich der WSI des ersten Winters um zehn Einheiten so sinkt die Wahrscheinlichkeit eines Luchsrisses um den Faktor $\exp(-0.102) = 0,90$
- Im Vergleich zu einem erwachsenen Tier sinkt die Wahrscheinlichkeit eines Luchsrisses um den Faktor $\exp(-0.629) = 0,53$

Der Wald, Score und Likelihood Ratio Test der globalen Null-Hypothese: $H_0: \boldsymbol{\beta} = \{\beta_1, \dots, \beta_p\} = \mathbf{0}$ wird ebenfalls im R-Output ausgegeben. Die Null-Hypothese für alle drei Tests chi-square mit p – Freiheitsgraden verteilt, sollte der p-Wert größer oder gleich 0,05 sein, so hat keine Kovariable einen signifikanten Einfluss (Therneau, et al., 2001).

Testergebnis für Modell mit festen Kovariablen:

Likelihood Ratio Test	=	38.61	5 df	$p \leq 0,0001$
Wald Test	=	32.27	5 df	$p \leq 0,0001$
Score Test	=	34.75	5 df	$p \leq 0,0001$

Tabelle 12: Null-Hypothese Tests für Modell mit festen Kovariablen

H_0 kann verworfen werden, mindestens ein Koeffizient hat einen signifikanten Einfluss

Als nächstes wird die **Interaktion** zwischen den Variablen berücksichtigt und das Modell nochmals mit sämtlichen, möglichen Interaktionen berechnet. Nach der Variablenselektion mittels der Funktion „step“ fallen jedoch sämtliche Interaktionen aus dem Modell somit bleibt das Haupteffektmodell wie oben beschrieben.

Die Berechnung des Einflusses des WSI ist sehr fraglich, denn schon in einigen biologischen Studien wurde nachgewiesen, dass die Rehe im Winter bei hoher Schneehöhe Schwierigkeiten haben sich fortzubewegen und so schlechter vor dem Luchs fliehen können (Andersen, et al., 1998). Dieses Problem könnte mit der Erweiterung des Cox-Modelles gelöst werden, denn hier wird der WSI für jeden Wintermonat neu berechnet und in das Modell eingefügt.

Modell 2f im Anhang

5.2.1 Residuen Modell 1f

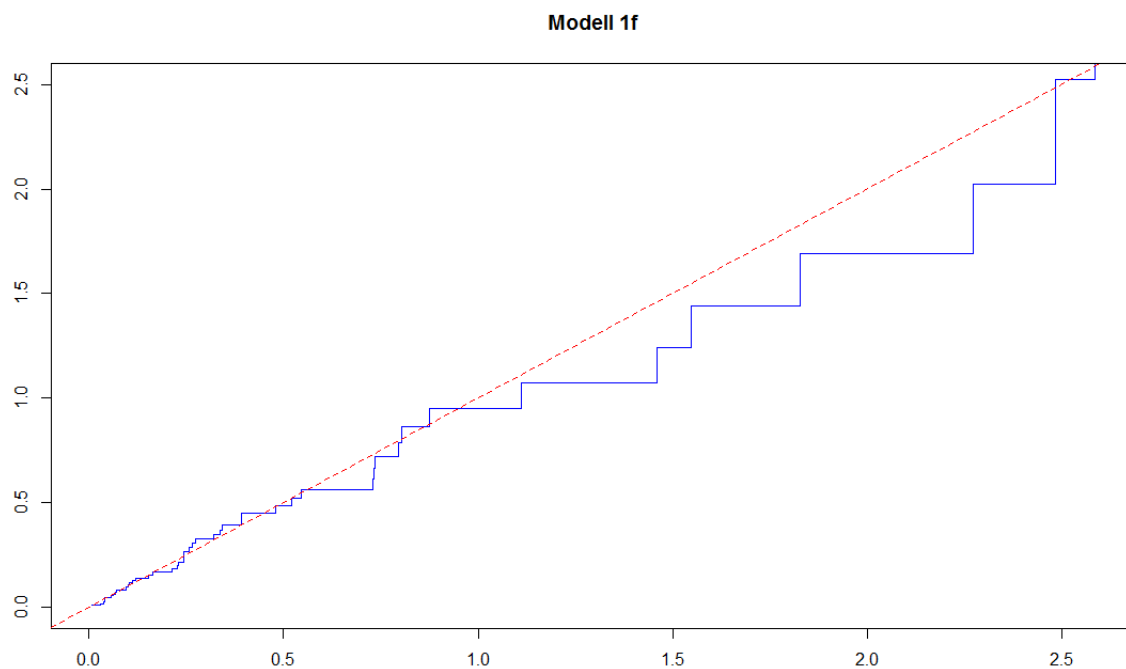


Abbildung 12: Cox-Snell Residuen von Modell 1f

In Abbildung 12 wurden die Cox-Snell-Residuen des Modell 1f mit festen Kovariablen geplottet. Der Nelson-Aalen-Schätzer weicht etwas von der 45° Linie ab. Insgesamt scheinen die Daten für das Modell geeignet zu sein.

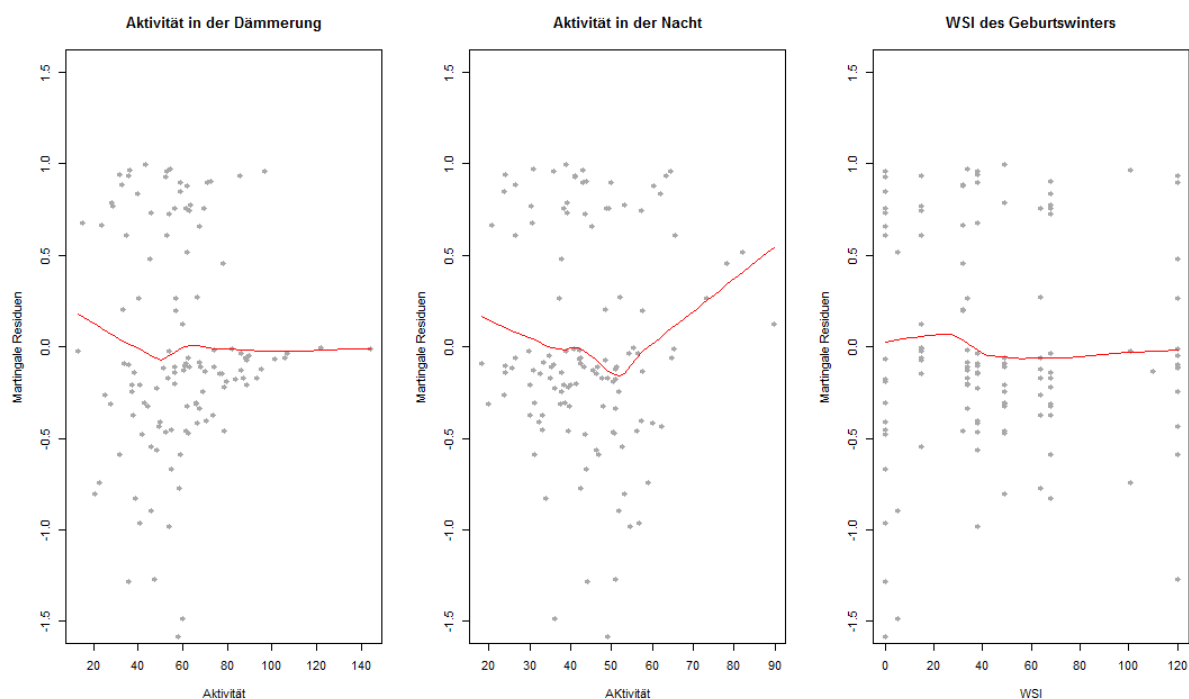


Abbildung 13: Martingale Residuen im Modell 1f

Bei den Martingalen-Residuen wird geprüft, ob eine funktionale Form der Variablen vorliegt. Dabei ist eine gerade Linie ein Zeichen, dass keine funktionale Form vorliegt und die Variable weiterhin so wie gehabt im Modell belassen werden kann. In der Abbildung 13 ist jedoch zu erkennen, dass besonders in der Variable „*Aktivität in der Nacht*“ eine funktionale Form vorzuliegen scheint. Allerdings gibt es nur wenige Beobachtungen, die zu dieser Annahme führen.

5.2.2 Test auf PH-Annahme

Wie schon in den Martingal-Residuen ersichtlich scheint die Variable „*Aktivität in der Nacht*“ eine funktionale Form aufzuweisen. Nun wird mit dem Grambsch und Therneau Test getestet, ob eine Verletzung der PH-Annahme besteht, in diesem Fall darf das Modell nicht angewendet werden.

Modell 1f	
Kovariablen	p
Aktivwert Dämmerung	0.4975
Aktivwert Nacht	0.0097
Geschlecht	0.5148
Alter	0.7589
WSI.Geburtswinter	0.4878
Global	0.1096

Tabelle 13: Grambsch und Therneau Test Modell 1f

Im Modell 1f ist eine Verletzung der PH-Annahme festgestellt worden, das Modell darf so nicht verwendet werden. Eine Möglichkeit die Variable trotzdem aufnehmen zu können wäre eine Schätzung der Variable „*Aktivität in der Nacht*“ über *psplines*:

5.2.3 Modell 1f mit pspline

Call:

```
coxph (formula = Surv(start, stop, event) ~ Alter + Geschlecht + Aktiv_daem
      + pspline(Aktiv_nacht, df=3) + WSI + Alter_Besenderung, data = Tabelle)
```

	Coef	exp(coef)	Pr(> z)
Aktiv Dämmerung	-0.0555	0.946	p < 0,0001
Pspline(Aktiv Nacht)	0.0270	0.0113	p < 0,0001
ps(aktivn)3		0.805	p < 0,0001
ps(aktivn)4		0.631	
ps(aktivn)5		0.518	
ps(aktivn)6		0.511	
ps(aktivn)7		0.632	
ps(aktivn)8		1.272	
ps(aktivn)9		2.260	
ps(aktivn)10		3.100	
ps(aktivn)11		3.642	
ps(aktivn)12		4.150	
Geschlecht männl.	1.0619	2.892	p < 0,0001
WSI Geburtswinter	-0.6957	0.989	p < 0,0001
Alter Jungtier	-0.0109	0.499	p < 0,0001

Tabelle 14: Modell 1f mit pspline nach Variablenselektion

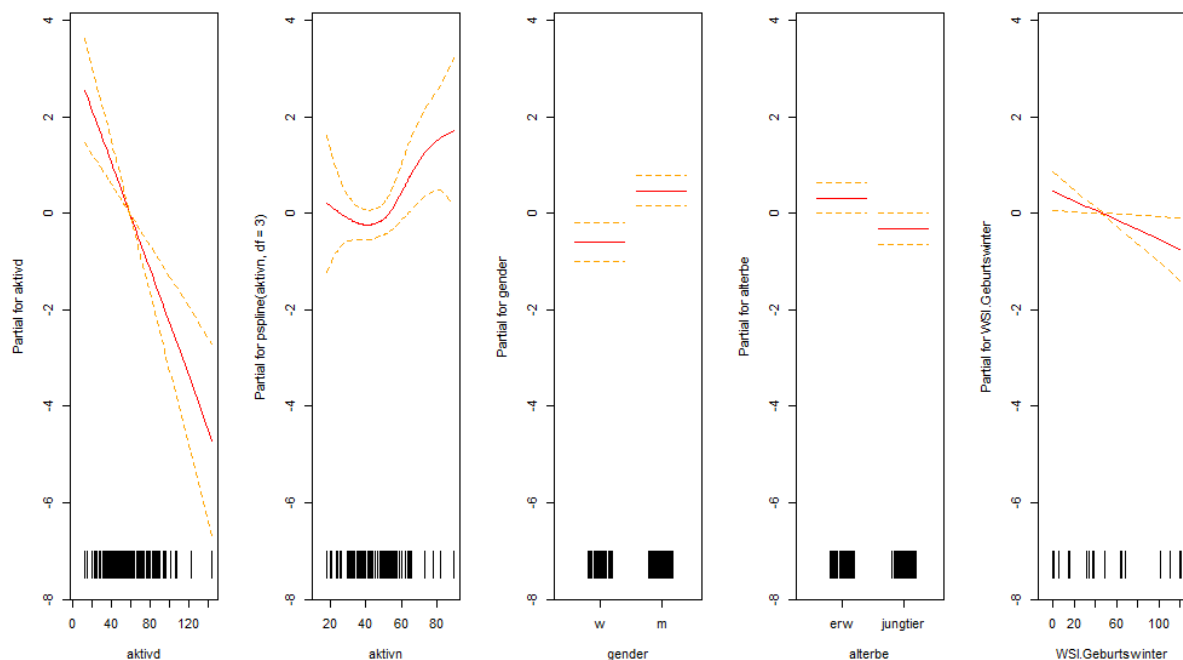


Abbildung 14: Partieller Einfluss mit pspline des Modells 1f

Die partiellen Einflüsse der anderen Kovariablen sind gleich geblieben. Der Koeffizient der Kovariable „Aktivität in der Nacht“ wurde nun mit psplines geschätzt und hat einen polynomen Verlauf.

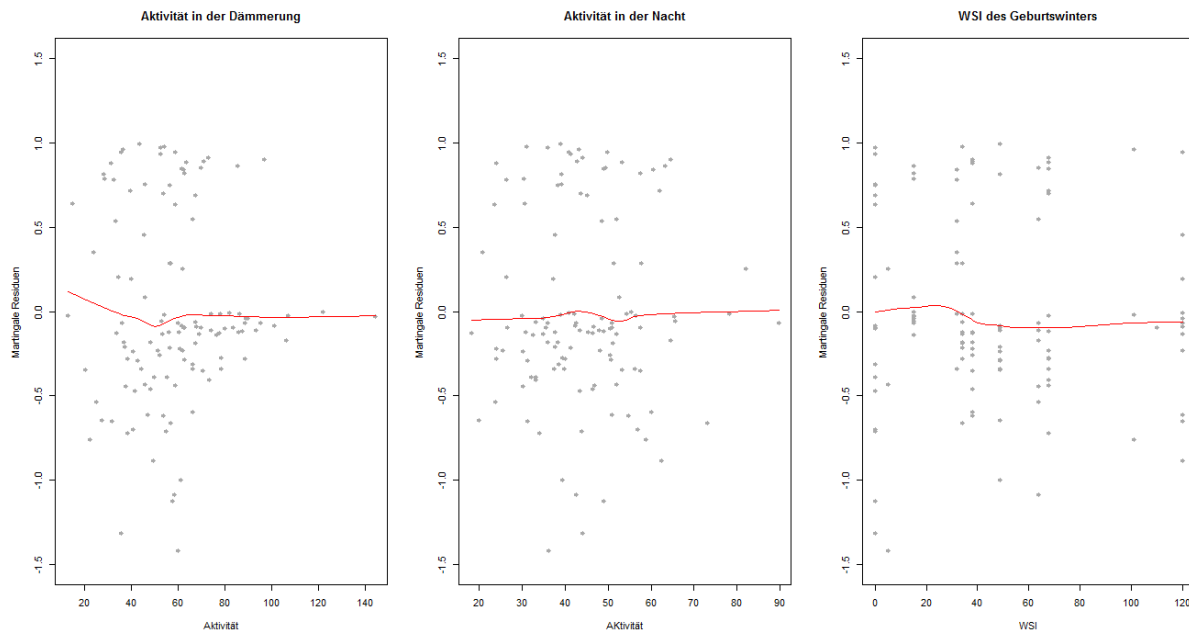


Abbildung 15: Martingale-Residuen von Modell 1f mit pspline

Nachdem die Variable mittels psplines aufgenommen wurde, ist die PH-Annahme nicht mehr verletzt und auch die Martingale-Residuen zeigen keine Verletzung der funktionalen Form. Da nur sehr wenige Beobachtungen mit sehr hoher Aktivität in der Nacht gemessen wurden, ist es fraglich, ob wirklich ein funktionaler Verlauf besser geeignet ist. Das Modell mit zeitvariierenden Kovariablen gibt hier einen sehr viel besseren Blick auf die Daten, da die Zahl der Beobachtungen durch den Zählprozess deutlich vergrößert wird.

5.3 Erweiterung des Cox-Modells mit zeitvariierenden Kovariablen

Das Cox-Modell mit festen Kovariablen gab bereits aufschlussreiche Informationen über die Risikofaktoren der Rehe, die zu einem Luchsriss führen können. Wie bereits im theoretischen Teil besprochen führt eine Erweiterung des Modells zu einer noch besseren Einschätzung der Einflussfaktoren über den gesamten Zeitraum. Die Untersuchung des Einflusses von Winter-Streng-Index auf das Risiko bzgl. eines Luchsrisses kann mit dem Modell mit festen Kovariablen nicht zufriedenstellend gelöst werden. Auch die auftretende, wechselnde Aktivität der Rehe wurde bis jetzt nur als ein Aktivwert aufgenommen, während in diesem Modell für jeden Monat eine Aktivwertbestimmung vorliegt, um den variierenden Einfluss dieser Variable aufzunehmen.

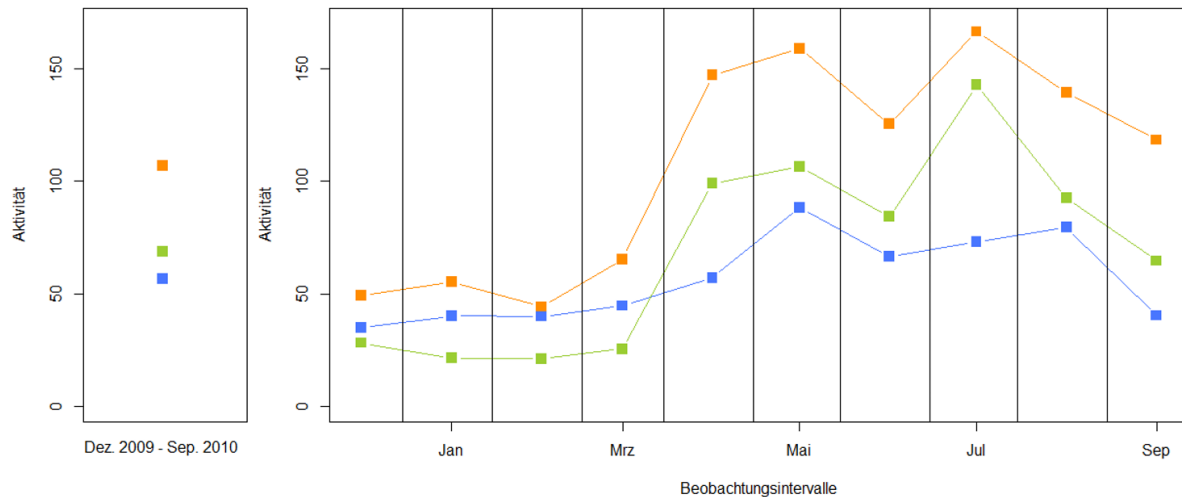


Abbildung 16: Gegenüberstellung der Variable „Aktivwert“ im Modell mit festen und mit variierenden Einflüssen

In Abbildung 16 ist auf der rechten Seite der Aktivverlauf, pro Monat, von Reh „Patrick“ geplottet und auf der linken Seite die gemittelten Werte in diesem Zeitraum dargestellt. Dabei sind die Werte der Dämmerung in orange, die Werte der Nacht in blau und die Werte des Tages in grün dargestellt.

Wie zu erkennen ist, wurde im Modell mit den festen Kovariablen (links) nur ein gemittelter Aktivwert pro Reh aufgenommen. Im Modell mit variierenden Kovariablen (rechts) wird je ein Aktivwert pro Beobachtungsintervall aufgenommen, dadurch werden die Daten erweitert. Das hat zur Folge, dass sich Änderungen in der Aktivität ins Modell aufnehmen lassen. Wenn z.B. ein Reh immer ein sehr aktives Verhalten gezeigt hat und im letzten Monat plötzlich dieses Verhalten ändert und vom Luchs gerissen wird, würde man diese Veränderung im Modell mit festen Kovariablen nicht bemerken, da hier der Aktivwert über den ganzen Zeitraum gemittelt wurde. Im Modell mit variierenden Kovariablen würde dieser Rückgang der Aktivität bemerkt werden und im Modell zur Aufklärung des Einflusses beitragen.

Durch die Erweiterung werden somit mehr Daten geschaffen, die einen größeren Aufschluss auf die Einflüsse geben können. Aus 120 Beobachtungen der Rehe werden im erweiterten Modell 1324 Beobachtungen.

5.4 Zeitvariierendes Cox-Modell mit Aktivwerten

In diesem Kapitel wird nun das Cox-Modell mit zeitvariierenden Kovariablen berechnet. Wie schon vorher besprochen werden Modell 1 und Modell 2 mit den Aktivwerten gefittet, während Modell 1a, 2a und 3a mit den Aktivanteilen berechnet werden.

5.4.1 Modell 1

Call:

```
coxph (formula = Surv(start, stop, event) ~ Alter + Geschlecht  
+ Aktiv_daem + Aktiv_nacht + Habitat + WSI, data = Tabelle)
```

Nach der Variablenselektion mittels der Funktion „step“ wird folgende Tabelle ausgegeben:

	Coef	exp(coef)	Pr(> z)
Aktiv Dämmerung	-0.036204	0.964444	p < 0,0001
Aktiv Nacht	0.017243	1.017393	0.0955
Geschlecht weibl	-0.597438	0.550219	0.0723
Alter	0.013247	1.013335	0.0635

Tabelle 15: Haupteffektmodell 1 nach Variablenselektion

Zur Visualisierung wurden die Einflüsse der Kovariablen mittels Grafikplots dargestellt.

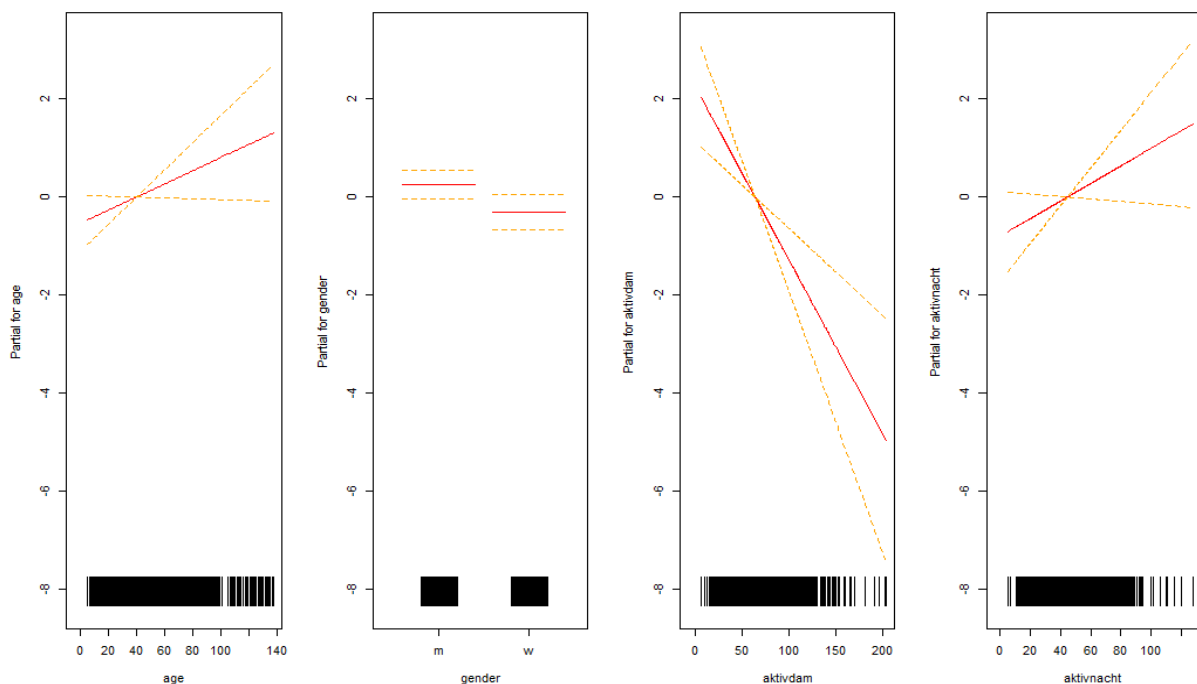


Abbildung 17: Einflüsse der Kovariablen im Modell 1

Nach der Variablenselektion mit der Funktion „step“ wurde die Kovariablen Winter-Streng-Index und Habitat aus dem Modell entfernt. Die verbleibenden Kovariablen sind Aktivwert

Dämmerung, Aktivwert Nacht, das Alter und das Geschlecht. Dabei sind die letzten drei genannten Variablen auf dem 10% Niveau signifikant.

Likelihood Ratio Test	=	22.76	4 df	p < 0,0001
Wald Test	=	19.35	4 df	p = 0,0005
Score Test	=	19.6	4 df	p = 0,0004

Tabelle 16: Null-Hypothese Tests für Modell 1

Interpretation:

- Wenn die Aktivität in der Dämmerung um **zehn** Einheiten ansteigt, so sinkt die Wahrscheinlichkeit eines Luchsrisses um den Faktor 0,7.
- Steigt die Aktivität dagegen in der Nacht um **zehn** Einheiten an, so steigt auch die Wahrscheinlichkeit eines Luchsrisses um den Faktor 1,19.
- Erhöht sich das Alter der Rehe um ein Jahr, steigt das Risiko eines Luchsrisses um den Faktor 1,17.
- Im Gegensatz zu einem männlichen Tier, sinkt die Wahrscheinlichkeit eines Luchsrisses für ein weibliches Tier um den Faktor 0,6.

Als nächstes wird die Interaktion zwischen den Variablen berücksichtigt und das Modell 1 nochmals mit sämtlichen, möglichen Interaktionen berechnet.

Modell 1 mit Interaktionen

Call:

```
coxph (formula = Surv(start, stop, event) ~ Alter + Geschlecht
      + Aktiv_daem + Aktiv_nacht
      + Alter:Geschlecht + Alter:Aktiv_daem + Alter:Aktiv_nacht
      + Geschlecht:Aktiv_daem + Geschlecht:Aktiv_nacht
      + Aktiv_daem:Aktiv_nacht , data = Tabelle)
```

Nach der Variablenselektion mittels der Funktion „step“ werden alle Interaktionen zwischen den Variablen aus dem Modell entfernt. Somit bleibt Modell 1 mit den Haupteffekten bestehen.

5.4.2 Modell 2

Call:

```
coxph (formula = Surv(start, stop, event) ~ Alter + Geschlecht
      + Aktiv_tag + Aktiv_nacht + Habitat + WSI, data = Tabelle)
```

Nach der Variablenselektion mittels der Funktion „step“ wird folgende Tabelle ausgegeben:

	coef	exp(coef)	Pr(> z)
Aktiv Tag	-0.053797	0.947625	0.0006
Aktiv Nacht	0.024714	1.025022	0.0252
WSI	0.027164	1.027536	0.0557
Alter	0.013397	1.013488	0.0512

Tabelle 17: Haupteffektemodell 2 nach Variablenselektion

Zur Visualisierung wurden die Einflüsse der Kovariablen mittels Grafikplots dargestellt.

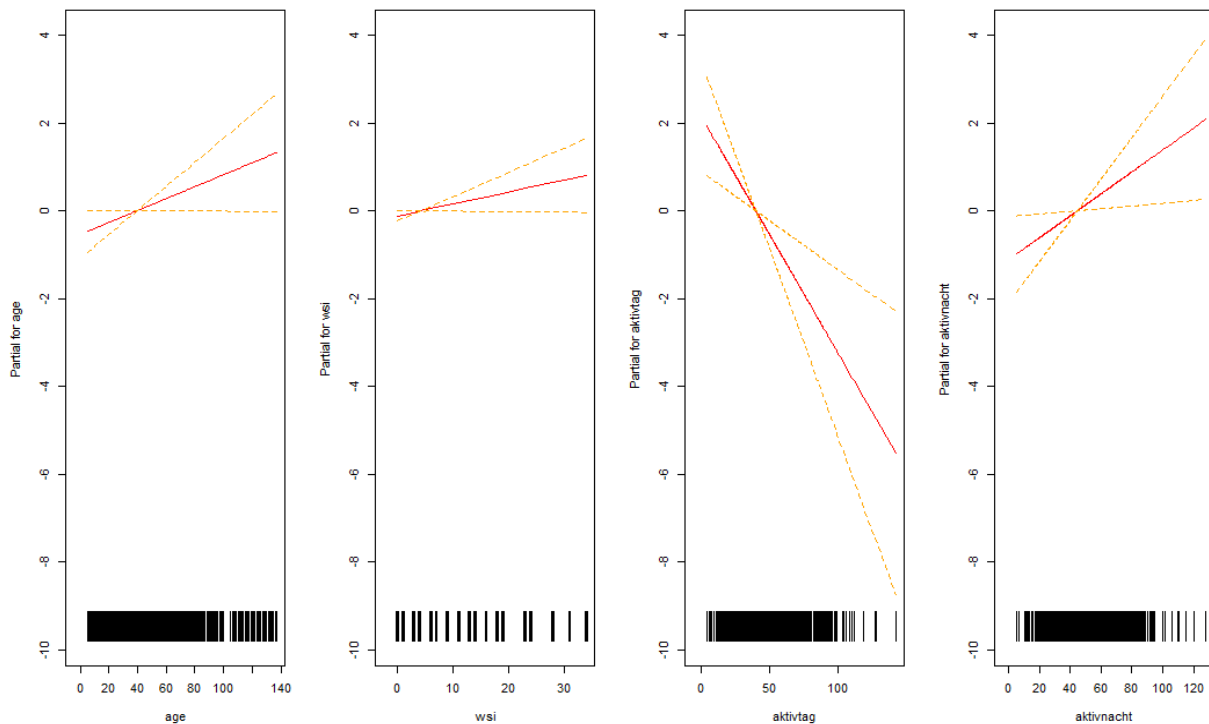


Abbildung 18: Einflüsse der Kovariablen im Modell 2

Nach der Variablenselektion wurde die Variable „Geschlecht“ aus dem Modell entfernt. Die Variable Winter-Streng-Index ist auf dem 10% Niveau signifikant.

Likelihood Ratio Test	=	22.16	4 df	p = 0,0002
Wald Test	=	19.43	4 df	p = 0,0006
Score Test	=	19.99	4 df	p = 0,0005

Tabelle 18: Null-Hypothese Tests für Modell 2

Interpretation:

- Steigt die Aktivität am Tag um **zehn** Einheiten so sinkt die Wahrscheinlichkeit eines Luchsrisses um den Faktor 0,58.
- Steigt die Aktivität in der Nacht um **zehn** Einheiten so erhöht sich das Risiko eines Luchsrisses um den Faktor 1,28.
- Erhöht sich das Alter der Rehe um ein Jahr, erhöht sich das Risiko eines Luchsrisses um den Faktor 1,18.
- Erhöht sich der Winter-Streng-Index um eine Einheit, so erhöht sich die Wahrscheinlichkeit eines Luchsrisses um den Faktor 1,03.

Wiederum wollen wir nun die Interaktionen zwischen den Variablen in einem Modell betrachten.

Modell 2 mit Interaktionen**Call:**

```
coxph (formula = Surv(start, stop, event) ~ Alter + WSI + Aktiv_tag  
      + Aktiv_nacht + Alter:WSI + Alter:Aktiv_tag + Alter:Aktiv_nacht  
      + WSI:Aktiv_daem + WSI:Aktiv_nacht + Aktiv_daem:Aktiv_nacht  
      , data = Tabelle)
```

Nach der Variablenselektion mittels der Funktion „step“ werden wiederum die Interaktionen aus dem Modell entfernt, somit bleibt das Haupteffektemodell 2 bestehen.

5.5 Zeitvariierendes Cox-Modell mit Aktivanteilen

Dadurch, dass die Aktivanteile der verschiedenen Tageszeiten stark miteinander korrelieren, werden in diesem Kapitel drei verschiedene Modelle berechnet.

5.5.1 Modell 1a

Call:

```
coxph (formula = Surv(start, stop, event) ~ Alter + Geschlecht  
+ Aktivanteil_daem + Habitat + WSI, data = Tabelle)
```

Nach der Variablenselektion mittels der Funktion „step“ wird folgende Tabelle ausgegeben:

	coef	exp(coef)	Pr(> z)
Aktivanteil Dämmerung	-0.044988	0.956009	p < 0,0001
Geschlecht weibl	-0.529551	0.588869	0.0911
Alter	0.011726	1.011795	0.0845

Tabelle 19: Haupteffektmodell 1a nach Variablenselektion

Wie bereits in den Kapitel mit den Aktivwerten wird wieder eine Grafik zur Visualisierung der Einflüsse in Abbildung x dargestellt.

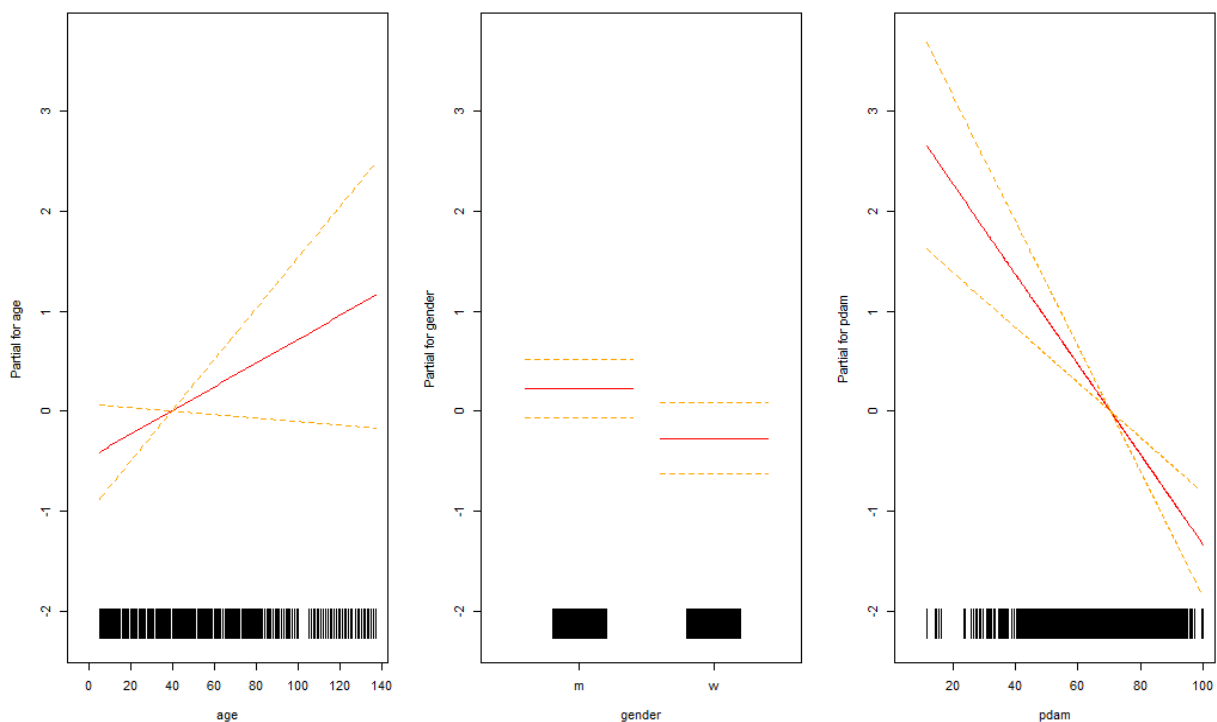


Abbildung 19: Einflüsse der Kovariablen im Modell 1a

Nach der Variablenselektion wurden nur die Variablen „Aktivanteil Dämmerung“, „Alter“ und „Geschlecht“ im Modell behalten. Variable „Geschlecht“ und „Alter“ sind auf dem 10% Niveau signifikant.

Likelihood Ratio Test	=	27.03	3 df	$p \leq 0,0001$
Wald Test	=	30.23	3 df	$p \leq 0,0001$
Score Test	=	31.26	3 df	$p \leq 0,0001$

Tabelle 20:Null-Hypothese Tests für Modell 1a

Interpretation:

- Steigt der Aktivitätsanteil eines Rehes in der Dämmerung um 10%, so sinkt das Risiko von einem Luchs gerissen zu werden um den Faktor 0,63.
- Erhöht sich das Alter der Rehe um ein Jahr, so steigt das Risiko eines Luchsrisses um den Faktor 1,15.
- Im Gegensatz zu einem männlichen Tier, sinkt das Risiko eines Luchsrisses für ein weibliches Tier um den Faktor 0,59.

Modell 1a mit Interaktionen

Call:

```
coxph (formula = Surv(start, stop, event) ~ Alter + Geschlecht+ Aktiv_daem
      + Alter:Geschlecht + Alter:Aktiv_daem + Geschlecht:Aktiv_daem
      , data = Tabelle)
```

Nach der Variablenselektion mittels der Funktion „step“ fallen alle Interaktionen aus dem Modell, das Modell mit den Haupteffekten bleibt.

5.5.2 Modell 2a

Call:

```
coxph (formula = Surv(start, stop, event) ~ Alter + Geschlecht
      + Aktivanteil_tag + Habitat + WSI, data = Tabelle)
```

Nach der Variablenselektion mittels der Funktion „step“ wird folgende Tabelle ausgegeben:

	coef	exp(coef)	Pr(> z)
Aktivanteil Tag	-0.01827	0.98189	0.0749
WSI	0.03123	1.03173	0.0273
Alter	0.01058	1.01063	0.1059

Tabelle 21: Haupteffektmodell 2a nach Variablenselektion

Wie bereits in den Kapiteln mit den Aktivwerten wird wieder eine Grafik zur Visualisierung der Einflüsse in Abbildung 15 dargestellt.

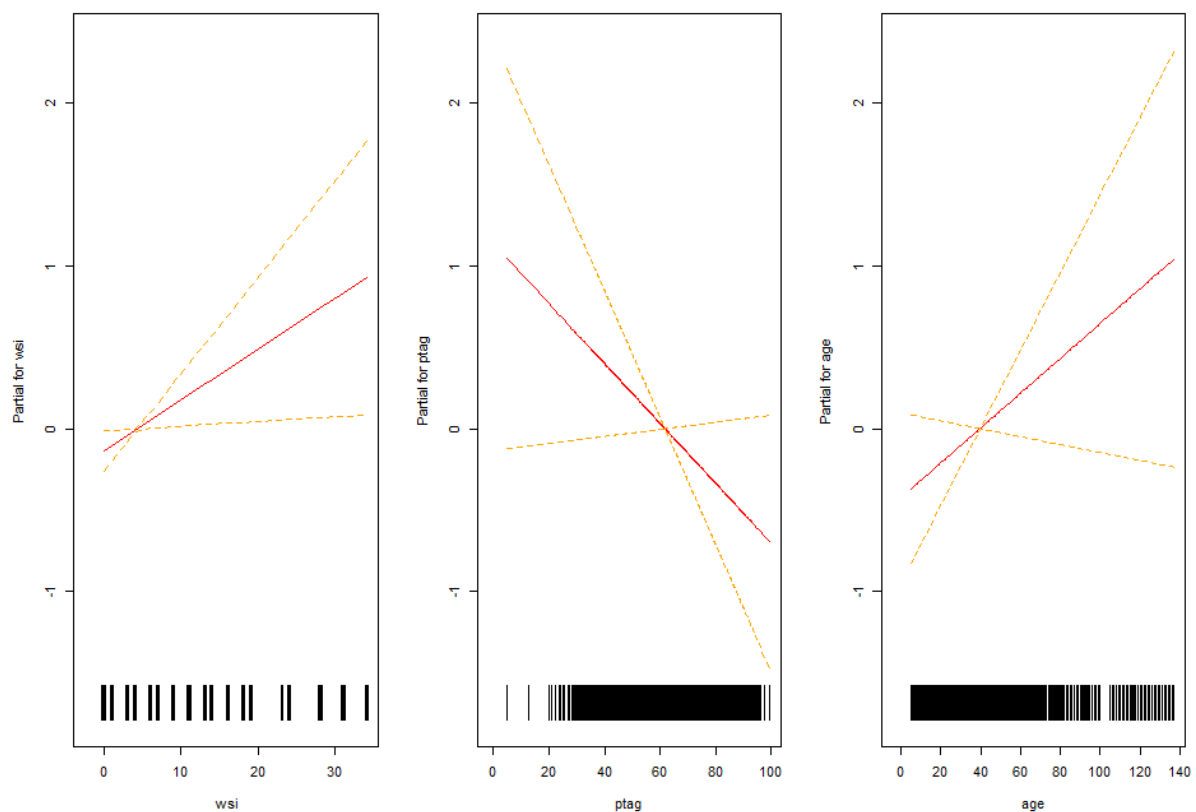


Abbildung 20: Einflüsse der Kovariablen im Modell 2a

Nach der Variablenselektion wurden nur die Variablen „Aktivanteil Tag“, „Alter“ und „WSI“ im Modell behalten

Likelihood Ratio Test	=	10.17	3 df	p = 0.01739
Wald Test	=	10.52	3 df	p = 0.01481
Score Test	=	10.82	3 df	p = 0.01288

Tabelle 22: Null-Hypothese Tests für Modell 2a

Interpretation:

- Steigt der Aktivitätsanteil eines Rehes am Tag um 10%, so sinkt das Risiko von einem Luchs gerissen zu werden um den Faktor 0,83.
- Erhöht sich das Alter der Rehe um ein Jahr, so steigt das Risiko eines Luchsrisses um den Faktor 1,45.
- Erhöht sich der WSI so steigt das Risiko eines Luchsrisses um den Faktor 1,03.

Modell 2a mit Interaktionen**Call:**

```
coxph (formula = Surv(start, stop, event) ~ Alter + WSI+ Aktiv_tag
      + Alter:WSI + Alter:Aktiv_tag + WSI:Aktiv_tag
      , data = Tabelle)
```

Nach der Variablenselektion mit der Funktion „step“ wird folgende Tabelle ausgegeben:

	coef	exp(coef)	Pr(> z)
Aktivanteil Tag	-0.0032457	0.9967595	0.7814
WSI	0.1678825	1.1827976	0.0008
Alter	0.0107870	1.0108454	0.1079
Aktivanteil Tag : WSI	-0.0026125	0.9973909	0.0082

Tabelle 23: Interaktionseffekte des Modells 2a

Die Interaktion zwischen Aktivanteil Tag und WSI ist als einzige Interaktion im Modell geblieben.

Likelihood Ratio Test	=	18.1	4 df	p = 0.0012
Wald Test	=	23.58	4 df	p ≤ 0,0001
Score Test	=	23.62	4 df	p ≤ 0,0001

Tabelle 24: Null-Hypothese Tests für Modell 2a mit Interaktionen

Interpretation der Interaktion:

- Wenn sich der WSI um **zehn** Einheiten erhöht, bei einem mittleren Aktivanteil am Tag von 62, so erhöht sich das Risiko eines Luchsrisses um den Faktor 1,06.
- Wenn sich der WSI um **zehn** Einheiten erhöht, bei einem hohen Aktivanteil am Tag von 90, so sinkt sich das Risiko eines Luchsrisses um den Faktor 0,51.
- Wenn sich der WSI um **zehn** Einheiten erhöht, bei einem niedrigen Aktivanteil am Tag von 30, so erhöht sich das Risiko eines Luchsrisses um den Faktor 1,09.

5.5.3 Modell 3a

Call:

```
coxph (formula = Surv(start, stop, event) ~ Alter + Geschlecht
      + Aktivanteil_nacht + Habitat + WSI, data = Tabelle)
```

Nach der Variablenselektion mittels der Funktion „step“ wird folgende Tabelle ausgegeben:

	coef	exp(coef)	Pr(> z)
Aktivanteil Nacht	-0.04114	0.95970	0.0003
WSI	0.02699	1.02736	0.0606
Alter	0.01237	1.01245	0.0574

Tabelle 25: Haupteffektmodell 3a nach Variablenselektion

Wie bereits in den Kapitel mit den Aktivwerten wird wieder eine Grafik zur Visualisierung der Einflüsse in Abbildung 16 dargestellt.

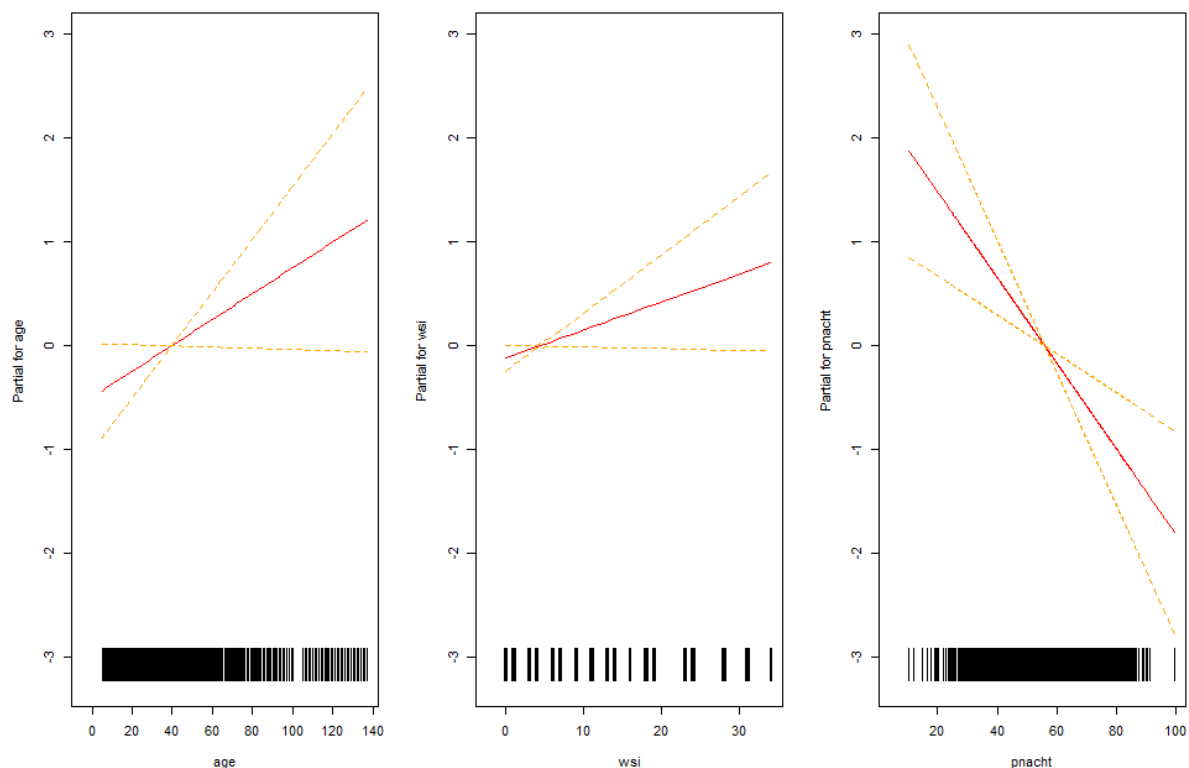


Abbildung 21: Einflüsse der Kovariablen in Modell 3a

Nach der Variablenselektion wurden nur die Variablen „Aktivanteil Nacht“, „Alter“ und „WSI“ im Modell behalten

Likelihood Ratio Test	=	21.05	3 df	p = 0.0001
Wald Test	=	21	3 df	p = 0,0001
Score Test	=	21.87	3 df	p ≤ 0,0001

Tabelle 26: Null-Hypothese Tests für Modell 3a

Interpretation:

- Steigt der Aktivitätsanteil eines Rehes in der Nacht um 10%, so sinkt das Risiko von einem Luchs gerissen zu werden um den Faktor 0,66.
- Erhöht sich das Alter der Rehe um ein Jahr, so steigt das Risiko eines Luchsrisses um den Faktor 1,16.
- Erhöht sich der WSI so steigt das Risiko eines Luchsrisses um den Faktor 1,03.

Modell 3a mit Interaktionen**Call:**

```
coxph (formula = Surv(start, stop, event) ~ Alter + WSI+ Aktiv_nacht
      + Alter:WSI + Alter:Aktiv_nacht + WSI:Aktiv_nacht
      , data = Tabelle)
```

Nach der Variablenselektion mit der Funktion „step“ wird folgende Tabelle ausgegeben:

	coef	exp(coef)	Pr(> z)
Aktivanteil Nacht	-0.0291526	0.9712682	0.0251
WSI	0.0915857	1.0959107	0.0183
Alter	0.0128737	1.0129569	0.0509
Aktivanteil Nacht : WSI	-0.0016824	0.9983190	0.0910

Tabelle 27: Interaktionseffekte des Modells 3a

Die Interaktion zwischen Aktivanteil Nacht und WSI ist als einzige Interaktion im Modell geblieben.

Likelihood Ratio Test	=	24.39	4 df	$p \leq 0,0001$
Wald Test	=	27.08	4 df	$p \leq 0,0001$
Score Test	=	28.79	4 df	$p \leq 0,0001$

Tabelle 28: Null-Hypothese Tests für Modell 3a mit Interaktionen

Interpretation:

- Wenn sich der WSI um **zehn** Einheiten erhöht, bei einem mittleren Aktivanteil in der Nacht von 56, so erhöht sich das Risiko eines Luchsrisses um den Faktor 1,02.
- Wenn sich der WSI um **zehn** Einheiten erhöht, bei einem hohem Aktivanteil in der Nacht von 85, so sinkt sich das Risiko eines Luchsrisses um den Faktor 0,59.
- Wenn sich der WSI um **zehn** Einheiten erhöht, bei einem niedrigen Aktivanteil in der Nacht von 30, so erhöht sich das Risiko eines Luchsrisses um den Faktor 1,49.
- Wenn sich die Aktivität in der Nacht um **zehn** Einheiten erhöht, bei einem WSI von 10, so sinkt das Risiko eines Luchsrisses um den Faktor 0,63.

5.6 Cox-Snell Residuen der Modelle mit variierenden Kovariablen

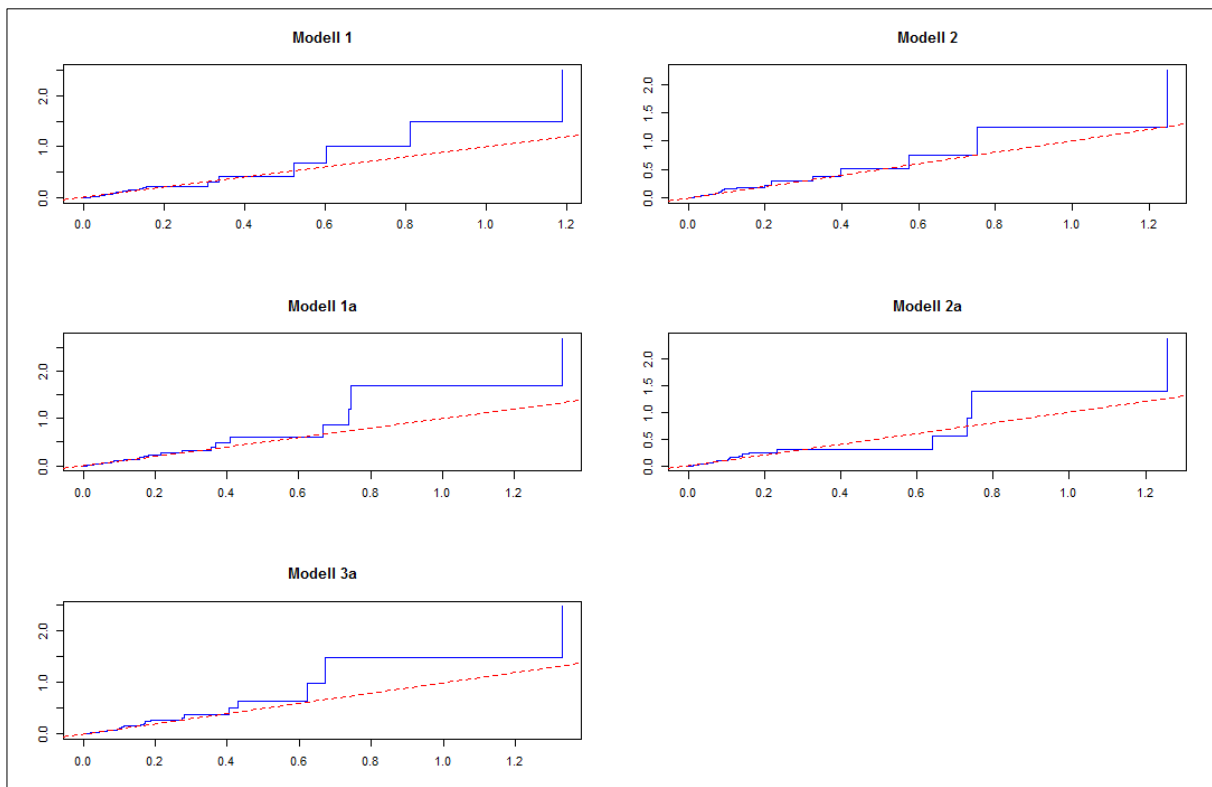


Abbildung 22: Cox-Snell-Residuen der fünf Cox-Modelle

In Abbildung 22 wurden die Cox-Snell-Residuen der fünf Modelle geplottet. In allen Modellen ist zu erkennen, dass der Nelson-Aalen-Schätzer am Ende abweicht. Dies kommt daher, dass am Ende der Zeit sehr wenige Beobachtungen stattgefunden haben. Insgesamt scheinen die Modelle, die mit den Aktivwerten arbeiten (Modell 1 und Modell 2) eine bessere Anpassung zu haben, als die Modelle, die mit den Aktivanteilen gerechnet worden sind. Gerade am Anfang der Berechnung, in der noch viele Individuen unter Risiko waren, ist die Eignung des Modells sehr gut. Alles in allem scheinen sich die Daten gut für die Methode des Cox-Modells zu eignen.

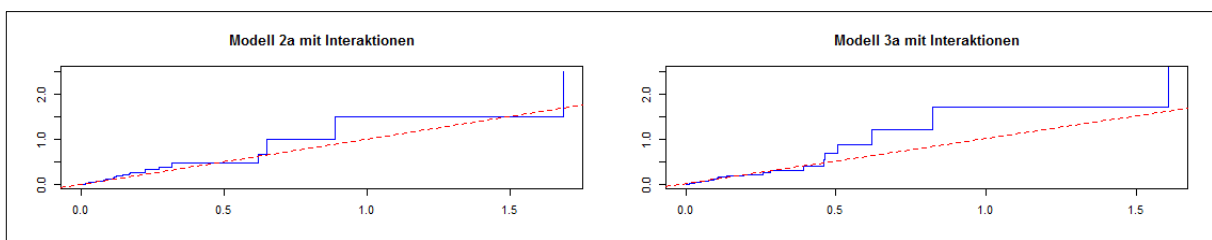


Abbildung 23: Cox-Snell-Residuen mit Interaktionen

In Abbildung 23 sind die Cox-Snell-Residuen der Modelle 2a und 3a mit Interaktionen geplottet. Auch hier ist keine grobe Verletzung der Eignung des Modells erkennbar.

5.7 Martingal-Residuen

Die Martingal-Residuen werden bei dem Modell mit variierenden Kovariablen per Beobachtung geplottet. Insgesamt konnte keine funktionale Form der Variablen festgestellt werden. Die Variablen müssen nicht geschichtet bzw. mit *psplines* geschätzt werden.

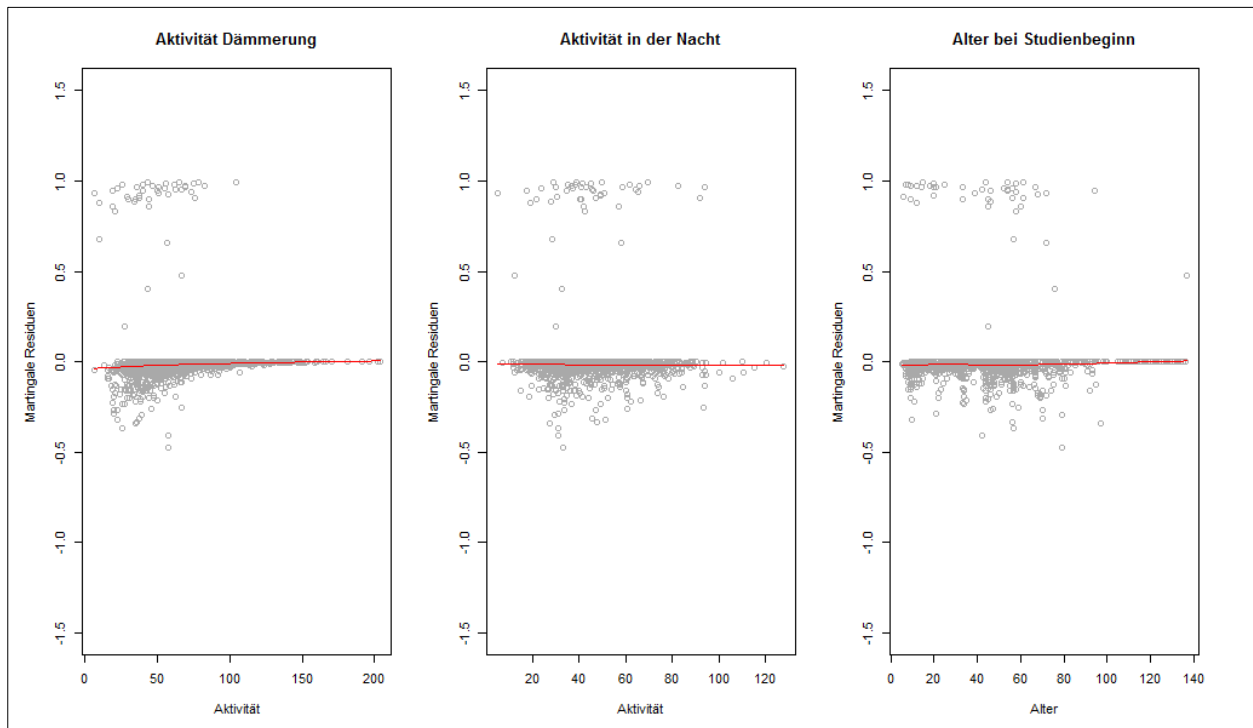


Abbildung 24: Martingal-Residuen Modell 1

In Abbildung 24 sind die Martingal-Residuen des ersten Modells dargestellt, die restlichen Abbildungen zu den Modellen befinden sich im Anhang. Keine der Kovariablen in den Modellen weist eine funktionale Form auf.

5.8 Test auf PH-Annahme

Modell 1	
Kovariablen	p
Aktivwert Dämmerung	0.812
Aktivwert Nacht	0.867
Geschlecht	0.722
Alter	0.874
Global	0.982

Modell 2	
Kovariablen	p
Aktivwert Tag	0.681
Aktivwert Nacht	0.991
WSI	0.261
Alter	0.875
Global	0.843

Modell 1a	
Kovariablen	p
Aktivanteil Dämmerung	0.341
Geschlecht	0.618
Alter	0.821
Global	0.780

Modell 2a	
Kovariablen	p
Aktivanteil Tag	0.895
WSI	0.373
Alter	0.886
Global	0.847

Modell 3a	
Kovariablen	p
Aktivanteil Nacht	0.352
WSI	0.504
Alter	0.568
Global	0.645

Modell 2a mit Interaktionen	
Kovariablen	p
Aktivanteil Tag	0.391
WSI	0.641
Alter	0.608
Aktivanteil Tag : WSI	0.536
Global	0.827

Modell 3a mit Interaktionen	
Kovariablen	p
Aktivanteil Nacht	0.180
WSI	0.613
Alter	0.623
Aktivanteil Nacht : WSI	0.558
Global	0.690

Tabelle 29: Grambsch und Therneau Test Ergebnisse der Modelle

Anders als im Modell mit festen Kovariablen ist hier keiner der p-Werte kleiner oder gleich 0.05, somit scheint bei allen zuvor gerechneten Modellen mit zeitvariierenden Kovariablen keine Verletzung der PH-Annahme vorzuliegen.

6. Fazit

In dieser Arbeit sollten die Risikofaktoren der Rehe bzgl. eines Luchsrisses analysiert werden. Die Risikofaktoren waren die Aktivität, der Aktivanteil, das Geschlecht, das Alter, das bevorzugte Habitat und der WSI. Bei den absoluten Aktivwerten zeigte sich in der deskriptiven Untersuchung, dass die Tiere in der Dämmerung eine deutlich höhere Aktivität zeigen als am Tag und in der Nacht. Auch schienen die untersuchten Tiere in der Nacht eine höhere Aktivität zu zeigen als am Tage. Der Aktivanteil zeigte wiederum, dass in der Dämmerung der höchste Aktivanteil gemessen wurde, allerdings hatten bei dieser Variable die Tiere einen höheren Aktivanteil am Tag als in der Nacht. In diesem Punkt unterscheiden sich die Aussagen zwischen Aktivwert und Aktivanteil.

Nachdem sieben verschiedene Modelle gerechnet und getestet wurden, kann folgende Aussage zur Datenanalyse getroffen werden:

- **Je höher die Aktivität in der Dämmerung und am Tag, desto geringer das Risiko eines Luchsrisses**
- **Je höher die Aktivität in der Nacht, desto höher ist das Risiko eines Luchsrisses**
- **Je höher der Aktivanteil in der Dämmerung, am Tag und in der Nacht, desto niedriger ist das Risiko eines Luchsrisses**
- **Weibliche Tiere haben ein geringeres Risiko vom Luchs gerissen zu werden als männliche Tiere**
- **Je älter das Tier, desto höher ist das Risiko eines Luchsrisses**
- **Je strenger der Winter, desto höher ist das Risiko eines Luchsrisses**
- **Das bevorzugte Habitat scheint keine Rolle bzgl. eines Luchsrisses zu spielen**

Dabei ist der Einfluss der Variablen „*Aktivität in der Dämmerung*“ und „*Aktivanteil Dämmerung*“ am höchsten. Insgesamt haben die Variablen der Aktivität einen viel deutlicheren Einfluss als die anderen Variablen.

Im Vergleich zu dem Modell mit festen Kovariablen bietet das Modell mit variierenden Kovariablen ein deutlicheres Bild von den Zusammenhängen. Die Variable Winter-Strenges-Index konnte in das Modell mit festen Kovariablen nicht zufriedenstellend aufgenommen werden, hier konnte nur der erste Winter berücksichtigt werden. Im Modell mit variierenden Kovariablen hingegen konnte diese Variable viel eindeutiger Ergebnisse liefern. Insgesamt konnten die Ergebnisse des Praktikums verifiziert werden.

Im Vergleich zu wenig aktiven Tieren haben aktivere Tiere ein geringeres Risiko bzgl. eines Luchsrisses. Damit wird die Hypothese, die im statistischen Praktikum gestellt wurde:

„Rehe mit einem höheren Aktivitätslevel haben eine höhere Wahrscheinlichkeit vom Luchs gerissen zu werden.“

durch die Datenanalyse in dieser Bachelorarbeit ebenfalls verworfen. (Zeis, et al., 2013)

Die zusätzliche Aufnahme der Variable „Aktivität in der Nacht“ war ebenfalls sehr aufschlussreich. Im Modell mit den festen Kovariablen zeigte sich durch wenige, hohe Werte der Nacht eine funktionale Form der Variable und somit eine PH-Annahmen-Verletzung. Durch die Berechnung mit zeitvariierenden Kovariablen konnte dies jedoch widerlegt werden, dort lag keine Verletzung der PH-Annahme für die Variable „Aktivität in der Nacht“ vor.

Das bevorzugte Habitat hatte keine Auswirkungen auf die Rehe bzgl. eines Luchsrisses, diese Variable wurde aus sämtlichen Modellen nach der Variablenselektion entfernt.

Durch diese Bachelorarbeit erhielt ich einen sehr interessanten Einblick in die Arbeit mit Überlebenszeitanalysen. Bereits im statistischen Praktikum wurde das erste Verständnis und Interesse für diese Art von Modellen geweckt. Mit der Analyse in dieser Bachelorarbeit wurde es weiter ausgebaut. Dabei stellte die Datenaufbereitung und Kodierung die größte Herausforderung dieser Untersuchung dar. Mit dem Statistikprogramm „R“ konnten die Analysen im Cox-Modell effizient berechnet werden. Die spannende Zusammenarbeit mit dem Projektpartner gab mir ebenfalls einen guten Einblick über die zukünftige Arbeit als Statistikerin.

Literaturverzeichnis

- Allison, Ph. D. Paul. 2012.** statisticalhorizons. [Online] Statistical Horizons, 2012.
<http://www.statisticalhorizons.com/>.
- Andersen, Reidar, Duncan, Patrick und D.C. Linnell, John. 1998.** The European Roe Deer: The Biology of Success. *The European Roe Deer: The Biology of Success*. Stockholm : Scandinavian University Press, 1998.
- Butz, Pius und Schulte, Robert. 2010.** vectronic-aerospace.com. [Online] VECTRONIC Aerospace GmbH, 2010. <http://www.vectronic-aerospace.com/>.
- Everitt, Brian und Hothorn, Thorsten. 2009.** *A Handbook of Statistical Analyses Using R*. Boca Raton, Florida : Chapman & Hall/CRC Press, 2009. ISBN 978-1-4200-7933-3.
- Fahrmeir, Ludwig. 2007.** statistik.lmu.de/~chris/survival/Skript07.pdf. [Online] 2007.
<http://www.statistik.lmu.de/~chris/survival/Skript07.pdf>.
- Galupki, Jürgen. 2012.** galupki.de/kalender/. [Online] 2012.
<http://galupki.de/kalender/?seite=sunmoon>.
- Höhle, Michael. 2009.** statistik.lmu.de. [Online] Statistik LMU, 2009.
http://www.statistik.lmu.de/institut/lehrstuhl/semwiso/lebensdaueranalyse_ws0809/download/regression_4.pdf.
- Klein, John und Moeschberger, Melvin. 1997.** *Survival Analysis; Techniques for Censored and Truncated Data*. New York : Springer-Verlag, 1997. ISBN 0-387-94829-5.
- Lawless, J.F. 1982.** *Statistical Models and Methods for Lifetime Data*. New York : John Wiley & Sons, 1982. ISBN 0-471-08544-8.
- Leibl, Dr. Franz. 2012.** nationalpark-bayerischer-wald.de. [Online] Nationalparkverwaltung Bayerischer Wald, 2012. <http://www.nationalpark-bayerischer-wald.de/>.
- Mayer, Verena. 2013 .** *Habitatwahl und Risiko für Rehe unter Luchsprädation (in Vorbereitung)*. München : s.n., 2013 .
- Potuschak, Heinrich. 2000.** *Lebensdaueranalysen mit zeitveränderlichen Kovariablen*. Linz : Universitätsverlag Rudolf Trauner, 2000. ISBN 3-85487-152-X.
- Rodriguez, German. 2007.** Generalized Linear Models. [Online] 2007.
<http://data.princeton.edu/wws509/notes/c7.pdf>.
- Therneau, Terry. 1996.** <http://www.mayo.edu/research>. *Technical Report Series: Extending the Cox Model*. [Online] 1996. <http://mayoresearch.mayo.edu/mayo/research/biostat/upload/58.pdf>.
- Therneau, Terry und Grambsch, Patricia. 2001.** *Modeling Survival Data - Extending the Cox Model*. New York : Springer Verlag, 2001. 0-387-98784-3.
- Wahlström, Kjell, D.C. Linnell, John und Gaillard, Jean-Michel. 1998.** From birth to independence: Birth growth, neonatal mortality, hiding behaviour und dispersal. [Buchverf.] Reidar Anderson und Patrick, D.C. Linnell, John Duncan. *The European Roe Deer: The Biology of Success*. Stockholm : Scandinavian University Press, 1998.
- Zeis, Klara, et al. 2013.** *Analyse des Einflusses zwischen individuellem Verhalten der Rehe und der Wahrscheinlichkeit eines Luchsrisses*. München : s.n., 2013.

Abbildungsverzeichnis

Tabelle 1: Erste Zeilen der Rohdaten des Rehes "Theresa" vom 16.12.2009	3
Tabelle 2: Ausschnitt aus den ersten Zeilen der sog "Kopfdaten"	4
Tabelle 3: Winter-Streng-Index der Klimastation Waldhäuser	10
Tabelle 4: Ausschnitt aus der Endtabelle nach Kodierung und Bereinigung	12
Tabelle 5: Kovariablen des Cox- Modells	30
Tabelle 6: Kovariablen des erweiterten Cox-Modells	30
Tabelle 7: VIF für Gesamtmodell, Modell 1f und Modell 2f mit festen Kovariablen	31
Tabelle 8: Varianz-Inflations-Faktor für Gesamtmodell, Modell 1 und Modell 2 mit Aktivwerten	32
Tabelle 9: Varianz-Inflations-Faktor für Gesamtmodell, Modell 1a und Modell 2a mit Aktivanteilen	32
Tabelle 10: Einflüsse der festen Kovariablen im Modell	34
Tabelle 11: Partieller Einfluss der Kovariablen im Modell 1f	34
Tabelle 12: Null-Hypothese Tests für Modell mit festen Kovariablen	35
Tabelle 13: Grambsch und Therneau Test Modell 1f	37
Tabelle 14: Modell 1f mit pspline nach Variablenselektion	38
Tabelle 16: Haupteffektemodell 1 nach Variablenselektion	41
Tabelle 17: Null-Hypothese Tests für Modell 1	42
Tabelle 18: Haupteffektemodell 2 nach Variablenselektion	43
Tabelle 19: Null-Hypothese Tests für Modell 2	43
Tabelle 20: Haupteffektemodell 1a nach Variablenselektion	45
Tabelle 21: Null-Hypothese Tests für Modell 1a	46
Tabelle 22: Haupteffektemodell 2a nach Variablenselektion	47
Tabelle 23: Null-Hypothese Tests für Modell 2a	48
Tabelle 24: Interaktionseffekte des Modells 2a	48
Tabelle 25: Null-Hypothese Tests für Modell 2a mit Interaktionen	48
Tabelle 26: Haupteffektemodell 3a nach Variablenselektion	49
Tabelle 27: Null-Hypothese Tests für Modell 3a	49
Tabelle 28: Interaktionseffekte des Modells 3a	50
Tabelle 29: Null-Hypothese Tests für Modell 3a mit Interaktionen	50
Tabelle 30: Grambsch und Therneau Test Ergebnisse der Modelle	53

Abbildung 1: Richtungsmessungen	3
Abbildung 2: Darstellung der Aufteilung der Tageszeiten im Jahresverlauf	6
Abbildung 3: Aktivwerte von Reh "David" am 07.03.2010	7
Abbildung 4: Fehlerhafte Software 2.0.8.....	8
Abbildung 5: Boxplots der Aktivverteilung in den Tageszeiten	13
Abbildung 6: Boxplots der Bewegungsanteile in den Tageszeiten	14
Abbildung 7: Schicksal der 120 untersuchten Rehe.....	15
Abbildung 8: Kaplan-Meier Überlebenskurven des Alters bzgl. eines Luchsrisses	16
Abbildung 9: Kaplan-Meier Überlebenskurve des Geschlechts bzgl. eines Luchsrisses.....	17
Abbildung 10: Darstellung der Aktivwerte bzgl. des Luchsrisses in den Tageszeiten	18
Abbildung 11: Darstellung der Aktivanteile bzgl. eines Luchsrisses in den Tageszeiten	19
Abbildung 12: Cox-Snell Residuen von Modell 1f	36
Abbildung 13: Martingale Residuen im Modell 1f	36
Abbildung 14: Partieller Einfluss mit pspline des Modells 1f	38
Abbildung 15: Martingale-Residuen von Modell 1f mit pspline	39
Abbildung 16: Gegenüberstellung der Variable „Aktivwert“ mit festen und variierenden Einflüssen ..	40
Abbildung 17: Einflüsse der Kovariablen im Modell 1	41
Abbildung 18: Einflüsse der Kovariablen im Modell 2	43
Abbildung 19: Einflüsse der Kovariablen im Modell 1a	45
Abbildung 20: Einflüsse der Kovariablen im Modell 2a	47
Abbildung 21: Einflüsse der Kovariablen in Modell 3a	49
Abbildung 22: Cox-Snell-Residuen der fünf Cox-Modelle.....	51
Abbildung 23: Cox-Snell-Residuen mit Interaktionen.....	51
Abbildung 24: Martingal-Residuen Modell 1.....	52
Abbildung 25: Korrelation zwischen den drei Aktivwerten	61
Abbildung 26: Geschlecht und Luchsriss	61
Abbildung 27: Beobachtungszeiträume der Rehe mit Lücken	62
Abbildung 28: Martingal-Residuen Modell 2.....	68
Abbildung 29: Martingal-Residuen Modell 1a	69
Abbildung 30: Martingal-Residuen 2a	69
Abbildung 31: Martingal-Residuen Modell 3a.....	70

Verwendete Software R:

R Core Team (2012). A Language and Environment for Statistical Computing. Vienna Austria. ISBN 3-900051-07-0

Survial-Package R:

Therneau T (2013). _A Package for Survival Analysis in S_. R package version 2.37-4, <URL: <http://CRAN.R-project.org/package=survival>>.

Terry M. Therneau and Patricia M. Grambsch (2000). _Modeling Survival Data: Extending the Cox Model_. Springer, New York. ISBN 0-387-98784-3.

Eidesstattliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Markt Schwaben, den 29.05.2013

(Klara Zeis)

Danksagung

Zunächst möchte ich mich bei meinen Geschwistern und Freunden bedanken, die mich während des Schreibens dieser Arbeit unterstützt und motiviert haben.

Ganz besonderen Dank möchte ich an Prof. Küchenhoff richten, der mich während der Anfertigung dieser Arbeit betreut hat. Vielen Dank für Ihre Geduld mit meinen vielen Fragen und für Ihre Mühen.

Des Weiteren möchte ich mich bei Herrn Dr. Heurich vom Nationalpark Bayerischer Wald für die gute Zusammenarbeit, den regen Mailkontakt und für die Möglichkeit meines Vortrages im Rahmen der Luchstagung bedanken.

Besonderen Dank an den kürzlich verstorbenen Prof. Leo Knüsel, ohne dessen Hilfe ich die ersten Semester des Studiums nicht so erfolgreich überstanden hätte.

Zuletzt möchte ich mich bei meinen Eltern bedanken, die mich nicht nur während des Studiums finanziell und emotional unterstützten, sondern auch zahlreiche Stunden Korrektur gelesen haben, was für Fachfremde eine große Herausforderung darstellte.

Anhang

Zusätzliche deskriptive Grafiken

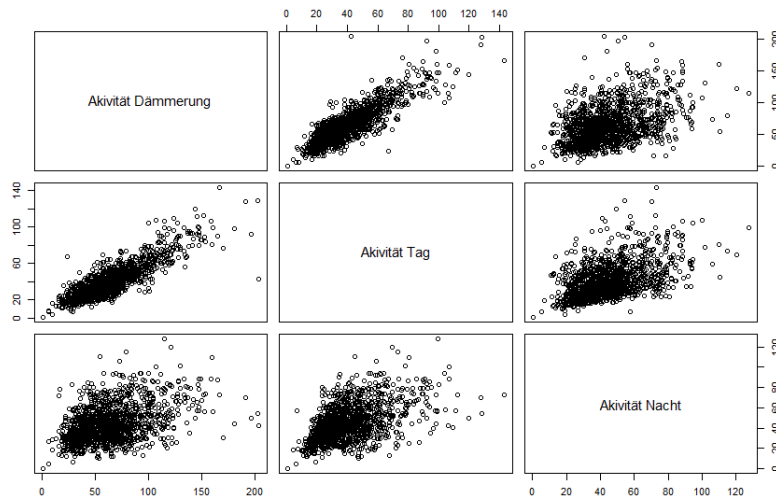


Abbildung 25: Korrelation zwischen den drei Aktivwerten

Abbildung 25 zeigt die Korrelation zwischen den drei Aktivwerten. Dabei kann man gut den starken linearen Zusammenhang zwischen den Aktivwerten am Tag und in der Dämmerung erkennen. Aufgrund dieses starken Zusammenhangs können die beiden Werte nicht in ein Modell aufgenommen werden.

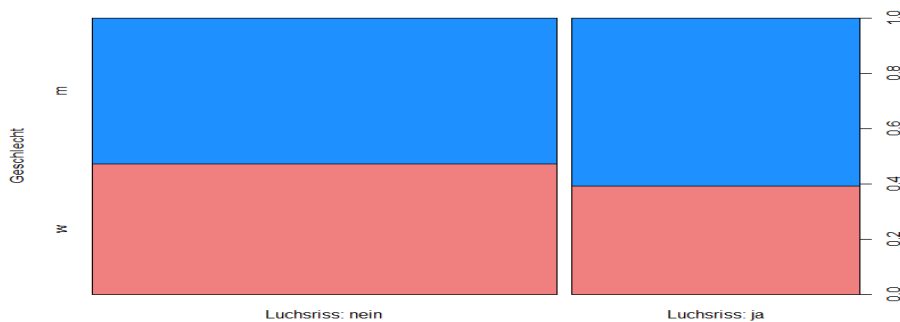


Abbildung 26: Geschlecht und Luchsriss

Abbildung 26 zeigt einen Mosaikplot, der den Zusammenhang zwischen dem Geschlecht bzgl. eines Luchsrisses darstellt. Auf der rechten Seite sind die 46 Tiere geplottet, die vom Luchs gerissen worden sind und auf der linken Seite die 74 Tiere, die nicht durch einen Luchsriss ums Leben gekommen sind. Die blaue Fläche gibt die Anzahl der männlichen Rehe wieder. Wie zu sehen ist, wurden etwas mehr männliche Tiere von dem Luchs gerissen, als weibliche Tiere. Der Unterschied ist jedoch sehr gering.



Abbildung 27: Beobachtungszeiträume der Rehe mit Lücken

Varianz-Inflations-Faktor

```
> #VIF der Modelle
>
> fit<- lm(event ~ pdam + pnacht + ptag + alter + gender + wsi + hab, fort)
> vif(fit)
      GVIF Df GVIFA(1/(2*Df))
pdam   3.865687 1      1.966135
pnacht 7.229199 1      2.688717
ptag   3.320317 1      1.822174
alter  1.078196 1      1.038362
gender 1.122567 1      1.059512
wsi    1.118867 1      1.057765
hab    1.217588 5      1.019882
> fit<- lm(event ~ pdam + alter + gender + wsi + hab, fort)
> vif(fit)
      GVIF Df GVIFA(1/(2*Df))
pdam   1.237897 1      1.112608
alter  1.033122 1      1.016426
gender 1.116640 1      1.056712
wsi    1.110550 1      1.053826
hab    1.161830 5      1.015113
> fit<- lm(event ~ ptag + alter + gender + wsi + hab, fort)
> vif(fit)
      GVIF Df GVIFA(1/(2*Df))
ptag   1.071992 1      1.035370
alter  1.048271 1      1.023851
gender 1.085552 1      1.041898
wsi    1.030834 1      1.015300
hab    1.092142 5      1.008853
> fit<- lm(event ~ pnacht + alter + gender + wsi + hab, fort)
> vif(fit)
      GVIF Df GVIFA(1/(2*Df))
pnacht 1.156444 1      1.075381
alter  1.033102 1      1.016416
gender 1.094236 1      1.046057
wsi    1.057111 1      1.028159
hab    1.158890 5      1.014855
>
>
> fit<- lm(event ~ aktivdam + aktivnacht + aktivtag + alter + gender + wsi + hab, fort)
> vif(fit)
      GVIF Df GVIFA(1/(2*Df))
aktivdam 4.286703 1      2.070436
aktivnacht 1.547583 1      1.244019
aktivtag 4.651319 1      2.156692
alter  1.043963 1      1.021745
gender 1.210764 1      1.100347
wsi    1.111257 1      1.054162
hab    1.309364 5      1.027321
> fit<- lm(event ~ aktivdam + aktivnacht + alter + gender + wsi + hab, fort)
> vif(fit)
      GVIF Df GVIFA(1/(2*Df))
aktivdam 1.600692 1      1.265185
aktivnacht 1.302884 1      1.141439
alter  1.041047 1      1.020317
gender 1.167387 1      1.080457
wsi    1.108576 1      1.052889
hab    1.264732 5      1.023764
> fit<- lm(event ~ aktivnacht + aktivtag + alter + gender + wsi + hab, fort)
> vif(fit)
      GVIF Df GVIFA(1/(2*Df))
aktivnacht 1.534078 1      1.238579
aktivtag 1.736842 1      1.317893
alter  1.043936 1      1.021732
gender 1.107496 1      1.052376
wsi    1.067712 1      1.033301
hab    1.265897 5      1.023858
```

R-Outputs der Modelle:

Modell 1f

Call:

```
coxph(formula = Surv(days, zensiert) ~ aktivd + aktivn + gender +
      WSI.Geburtswinter + alterbe, data = finalefix)
```

n= 120, number of events= 46

	coef	exp(coef)	se(coef)	z	Pr(> z)	
aktivd	-0.056808	0.944775	0.011424	-4.973	6.6e-07	***
aktivn	0.026553	1.026909	0.013001	2.042	0.04111	*
genderm	1.042920	2.837490	0.355165	2.936	0.00332	**
WSI.Geburtswinter	-0.010225	0.989827	0.004269	-2.395	0.01662	*
alterbejungtier	-0.628854	0.533203	0.318220	-1.976	0.04814	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
aktivd	0.9448	1.0585	0.9239	0.9662
aktivn	1.0269	0.9738	1.0011	1.0534
genderm	2.8375	0.3524	1.4145	5.6918
WSI.Geburtswinter	0.9898	1.0103	0.9816	0.9981
alterbejungtier	0.5332	1.8755	0.2858	0.9949

Concordance= 0.752 (se = 0.049)

Rsquare= 0.275 (max possible= 0.952)

Likelihood ratio test= 38.61 on 5 df, p=2.849e-07

Wald test = 32.27 on 5 df, p=5.259e-06

Score (logrank) test = 34.75 on 5 df, p=1.689e-06

Modell 2f

Call:

```
coxph(formula = Surv(days, zensiert) ~ aktivt + alterbe + gender +
      WSI.Geburtswinter + aktivn, data = finalefix)
```

n= 120, number of events= 46

	coef	exp(coef)	se(coef)	z	Pr(> z)	
aktivt	-0.112626	0.893485	0.022256	-5.060	4.18e-07	***
alterbejungtier	-0.648627	0.522763	0.315093	-2.059	0.03954	*
genderm	0.628183	1.874202	0.340949	1.842	0.06541	.
WSI.Geburtswinter	-0.009698	0.990348	0.004372	-2.218	0.02654	*
aktivn	0.047643	1.048796	0.014547	3.275	0.00106	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
aktivt	0.8935	1.1192	0.8553	0.9333
alterbejungtier	0.5228	1.9129	0.2819	0.9694
genderm	1.8742	0.5336	0.9607	3.6562
WSI.Geburtswinter	0.9903	1.0097	0.9819	0.9989
aktivn	1.0488	0.9535	1.0193	1.0791

Concordance= 0.776 (se = 0.049)

Rsquare= 0.303 (max possible= 0.952)

Likelihood ratio test= 43.24 on 5 df, p=3.304e-08

Wald test = 32.95 on 5 df, p=3.848e-06

Score (logrank) test = 37.42 on 5 df, p=4.928e-07

Modell mit zeitvariierenden Kovariablen:

Modell 1

call:

```
coxph(formula = Surv(startstudy, stopstudy, event) ~ alter +
      gender + aktivdam + aktivnacht, data = f)
```

n= 1324, number of events= 46

	coef	exp(coef)	se(coef)	z	Pr(> z)
alter	0.013247	1.013335	0.007139	1.855	0.0635 .
genderw	-0.597438	0.550219	0.332418	-1.797	0.0723 .
aktivdam	-0.036204	0.964444	0.008888	-4.073	4.64e-05 ***
aktivnacht	0.017243	1.017393	0.010344	1.667	0.0955 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
alter	1.0133	0.9868	0.9993	1.0276
genderw	0.5502	1.8175	0.2868	1.0556
aktivdam	0.9644	1.0369	0.9478	0.9814
aktivnacht	1.0174	0.9829	0.9970	1.0382

Concordance= 0.709 (se = 0.052)
 Rsquare= 0.018 (max possible= 0.237)
 Likelihood ratio test= 23.54 on 4 df, p=9.858e-05
 Wald test = 20.06 on 4 df, p=0.0004861
 Score (logrank) test = 20.28 on 4 df, p=0.0004398

Modell 2

call:

```
coxph(formula = Surv(startstudy, stopstudy, event) ~ age + wsi +
      aktivtag + aktivnacht, data = f)
```

n= 1324, number of events= 46

	coef	exp(coef)	se(coef)	z	Pr(> z)
age	0.013397	1.013488	0.006872	1.949	0.051242 .
wsi	0.027164	1.027536	0.014197	1.913	0.055698 .
aktivtag	-0.053797	0.947625	0.015685	-3.430	0.000604 ***
aktivnacht	0.024714	1.025022	0.011044	2.238	0.025234 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
age	1.0135	0.9867	0.9999	1.0272
wsi	1.0275	0.9732	0.9993	1.0565
aktivtag	0.9476	1.0553	0.9189	0.9772
aktivnacht	1.0250	0.9756	1.0031	1.0475

Concordance= 0.7 (se = 0.052)
 Rsquare= 0.017 (max possible= 0.237)
 Likelihood ratio test= 22.19 on 4 df, p=0.0001838
 Wald test = 19.44 on 4 df, p=0.0006449
 Score (logrank) test = 19.98 on 4 df, p=0.0005034

Modell 1a

```
Call:
coxph(formula = Surv(startstudy, stopstudy, event) ~ age + gender +
      pdam, data = f)
```

```
n= 1324, number of events= 46
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
age	0.011726	1.011795	0.006796	1.725	0.0845 .
genderw	-0.529551	0.588869	0.323015	-1.639	0.1011
pdam	-0.044988	0.956009	0.008690	-5.177	2.26e-07 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
age	1.0118	0.9883	0.9984	1.0254
genderw	0.5889	1.6982	0.3127	1.1091
pdam	0.9560	1.0460	0.9399	0.9724

```
Concordance= 0.666 (se = 0.052 )
Rsquare= 0.02 (max possible= 0.237 )
Likelihood ratio test= 27.42 on 3 df, p=4.802e-06
Wald test = 30.75 on 3 df, p=9.598e-07
Score (logrank) test = 31.92 on 3 df, p=5.434e-07
```

Modell 2a

```
Call:
coxph(formula = Surv(startstudy, stopstudy, event) ~ wsi + ptag +
      age, data = f)
```

```
n= 1324, number of events= 46
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
wsi	0.03123	1.03173	0.01415	2.207	0.0273 *
ptag	-0.01827	0.98189	0.01026	-1.781	0.0749 .
age	0.01058	1.01063	0.00654	1.617	0.1059

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
wsi	1.0317	0.9692	1.0035	1.061
ptag	0.9819	1.0184	0.9623	1.002
age	1.0106	0.9895	0.9978	1.024

```
Concordance= 0.601 (se = 0.052 )
Rsquare= 0.008 (max possible= 0.237 )
Likelihood ratio test= 10.14 on 3 df, p=0.01739
Wald test = 10.49 on 3 df, p=0.01481
Score (logrank) test = 10.8 on 3 df, p=0.01288
```

Modell 2a mit Interaktionen

```
Call:
coxph(formula = Surv(startstudy, stopstudy, event) ~ age + wsi +
      ptag + wsi:ptag, data = f)
```

```
n= 1340, number of events= 46
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
age	0.0107870	1.0108454	0.0067109	1.607	0.107968
wsi	0.1678825	1.1827976	0.0500550	3.354	0.000797 ***
ptag	-0.0032457	0.9967595	0.0116944	-0.278	0.781361
wsi:ptag	-0.0026125	0.9973909	0.0009877	-2.645	0.008167 **

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	exp(coef)	exp(-coef)	lower .95	upper .95
age	1.0108	0.9893	0.9976	1.0242
wsi	1.1828	0.8455	1.0723	1.3047
ptag	0.9968	1.0033	0.9742	1.0199
wsi:ptag	0.9974	1.0026	0.9955	0.9993

```
Concordance= 0.662 (se = 0.052 )
Rsquare= 0.013 (max possible= 0.235 )
Likelihood ratio test= 18.1 on 4 df, p=0.001179
Wald test = 23.58 on 4 df, p=9.69e-05
Score (logrank) test = 23.62 on 4 df, p=9.501e-05
```

Modell 3a

```
Call:
coxph(formula = Surv(startstudy, stopstudy, event) ~ age + wsi +
      pnacht, data = f)
```

```
n= 1324, number of events= 46
```

	coef	exp(coef)	se(coef)	z	Pr(> z)
age	0.01237	1.01245	0.00651	1.900	0.057463 .
wsi	0.02699	1.02736	0.01438	1.877	0.060565 .
pnacht	-0.04114	0.95970	0.01126	-3.654	0.000258 ***

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	exp(coef)	exp(-coef)	lower .95	upper .95
age	1.0124	0.9877	0.9996	1.0254
wsi	1.0274	0.9734	0.9988	1.0567
pnacht	0.9597	1.0420	0.9388	0.9811

```
Concordance= 0.642 (se = 0.052 )
Rsquare= 0.016 (max possible= 0.237 )
Likelihood ratio test= 20.99 on 3 df, p=0.0001059
Wald test = 20.94 on 3 df, p=0.0001083
Score (logrank) test = 21.8 on 3 df, p=7.167e-05
```

Modell 3a mit Interaktionen

```
Call:
coxph(formula = Surv(startstudy, stopstudy, event) ~ age + wsi +
      pnacht + wsi:pnacht, data = f)
```

n= 1324, number of events= 46

	coef	exp(coef)	se(coef)	z	Pr(> z)
age	0.0128737	1.0129569	0.0065928	1.953	0.0509 .
wsi	0.0915857	1.0959107	0.0388171	2.359	0.0183 *
pnacht	-0.0291526	0.9712682	0.0130125	-2.240	0.0251 *
wsi:pnacht	-0.0016824	0.9983190	0.0009953	-1.690	0.0910 .

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

	exp(coef)	exp(-coef)	lower .95	upper .95
age	1.0130	0.9872	1.0000	1.0261
wsi	1.0959	0.9125	1.0156	1.1825
pnacht	0.9713	1.0296	0.9468	0.9964
wsi:pnacht	0.9983	1.0017	0.9964	1.0003

Concordance= 0.651 (se = 0.052)
 Rsquare= 0.018 (max possible= 0.237)
 Likelihood ratio test= 24.24 on 4 df, p=7.139e-05
 Wald test = 26.91 on 4 df, p=2.077e-05
 Score (logrank) test = 28.59 on 4 df, p=9.484e-06

Martingal-Residuen der Modelle

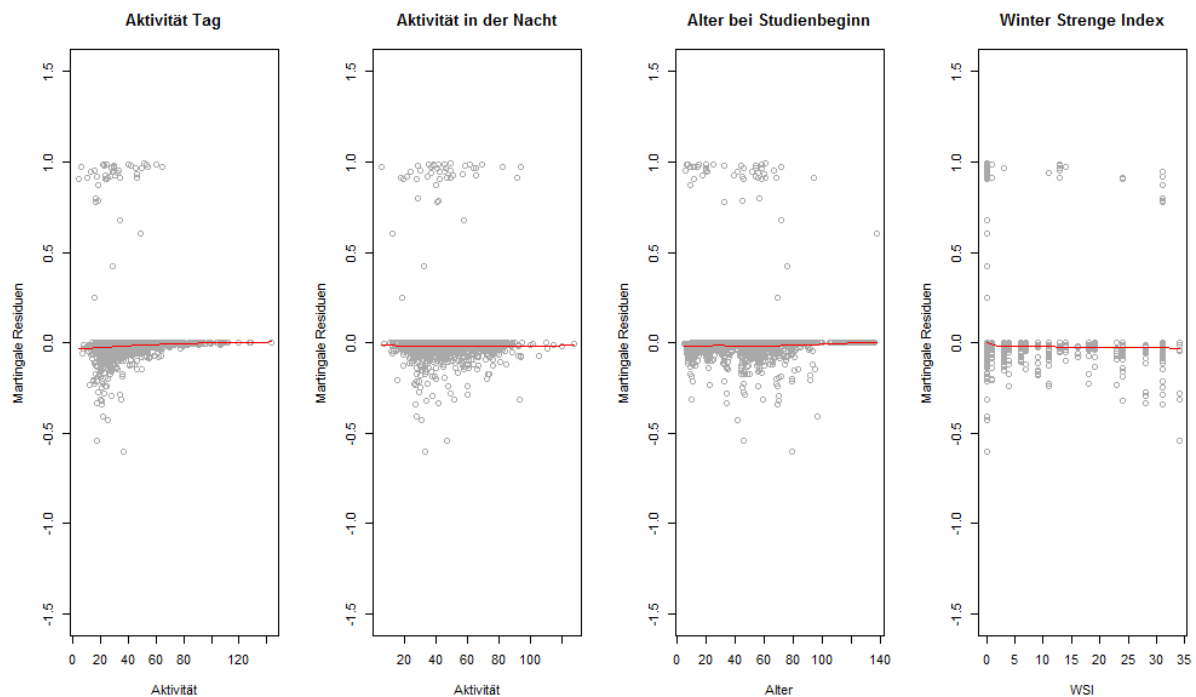


Abbildung 28: Martingal-Residuen Modell 2

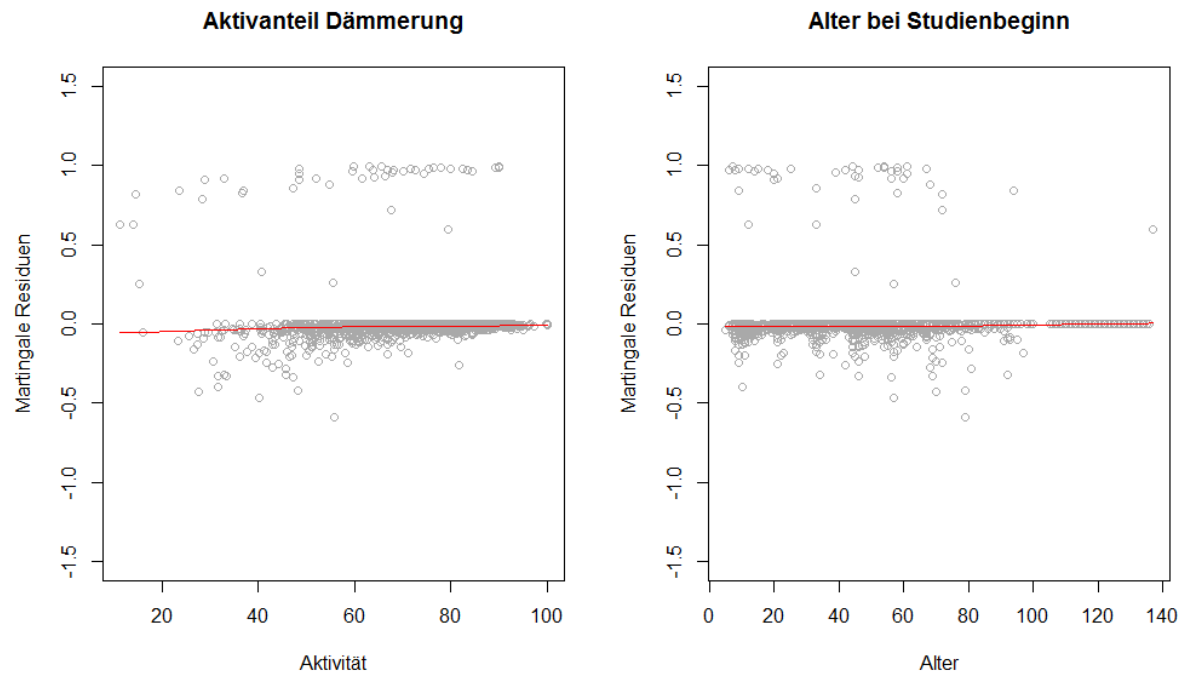


Abbildung 29: Martingale-Residuen Modell 1a

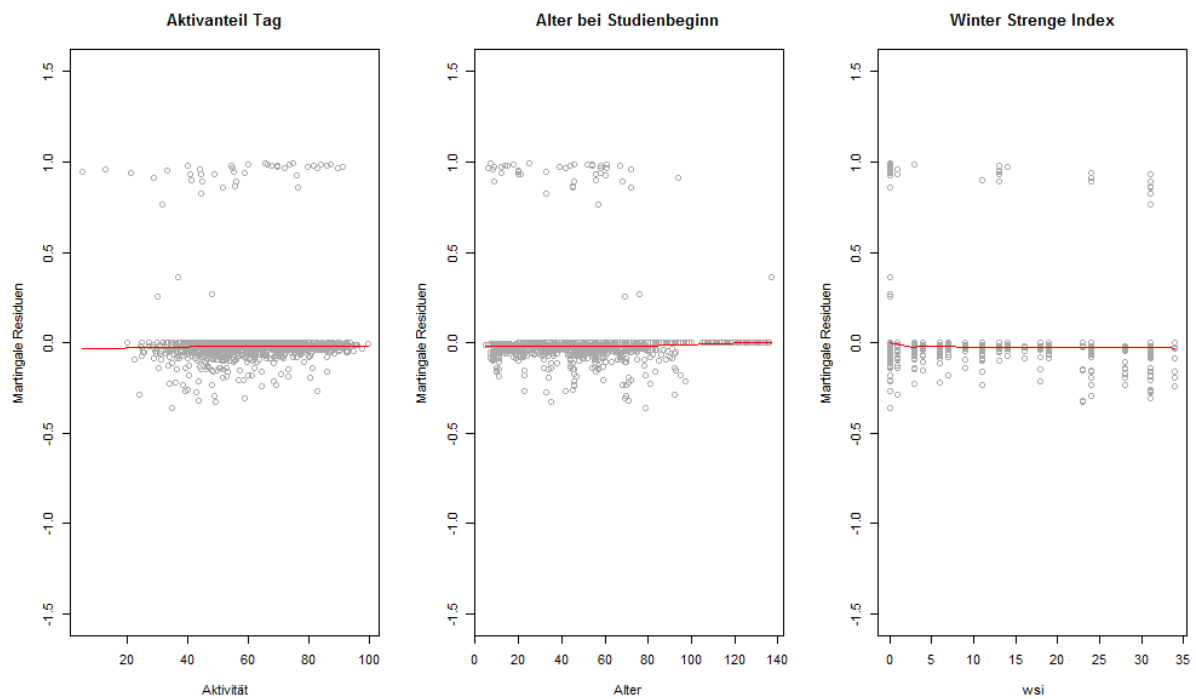


Abbildung 30: Martingale-Residuen 2a

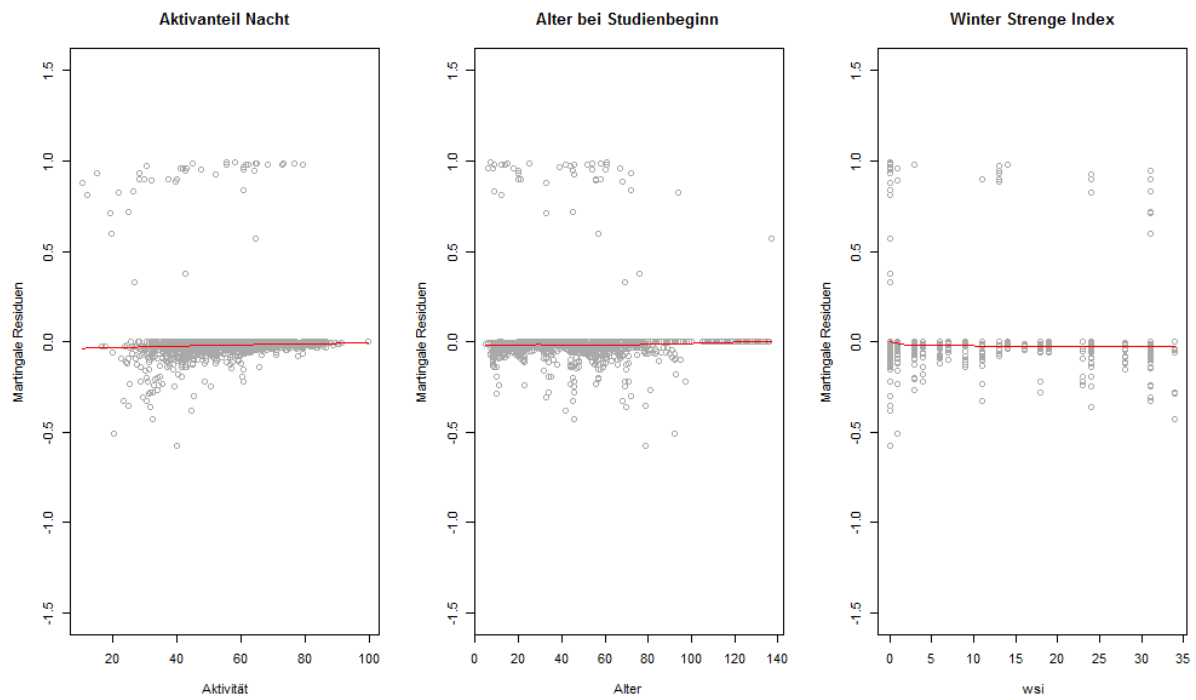


Abbildung 31: Martingale-Residuen Modell 3a